Essays in Economic History and Development

Thesis advisors	Author
J. Robinson, N. Nunn and R. Hornbeck	**P. Osafo-Kwaako**

Essays in Economic History and Development

Abstract

Chapter 1 provides a brief overview of the recent literature in economic history and long-run development, and summarizes the main findings of the three essays presented in this dissertation.

In Chapter 2, I examine the subject of villagization in Tanzania, a major episode of development planning in post-independence Tanzania. I revisit this period of Tanzania's economic history, focusing on the legacy of developmental villages (*vijiji uya maendeleo*) introduced in mainland Tanzania over the period 1974-1982. Combining historical data on Tanzania from the 1970s with data from population censuses and recent national household surveys, I investigate whether variation in the intensity of the governments villagization program explains within-region variation in social and economic outcomes today. I document that, in the short-run, developmental villages led to an increase in various educational outcomes, such as primary school completion rates, literacy rates, and total years of schooling. Today, districts which experienced a high share of developmental villages have greater availability of some public goods and citizens report higher rates of participation in community activities, but there is worse perception of corruption among government officials and greater rejection of one-party rule. Per capita household consumption is also significantly *lower* in districts with historically high levels of the treatment measure. To address potential endogeneity in village formation, I report instrumental variable results based on variation in ethnolinguistic fragmentation and the occurrence of droughts in the 1970s which facilitated the resettlement of peasants into villages. I conclude by providing some preliminary evidence on the lack of economic diversification as well as political alignment to the TANU/CCM party as possible channels which explain the legacy of the villagization experiment.

In Chapter 3, I turn to the subject of disease eradication, and examine the impact of the successful control of a highly infectious tropical disease, yaws, in Ghana over the period

1956-1963. The availability of cheap, mass-produced penicillin following World War II resulted in a mass treatment campaign by WHO/UNICEF aimed at controlling the prevalence of yaws and other bacterial infections. I examine the effect of this penicillin campaign in which over 70 percent of the estimated Ghanaian population received a single dose of an intramuscular penicillin injection. Data collected by the WHO/UNICEF program before and after the campaign indicates that penicillin-based treatment resulted in an immediate reduction in the prevalence of infectious yaws among the Ghanaian population. Using a microsample from the 2000 Ghanaian census, I estimate a difference-in-difference model exploiting spatial variation in pre-treatment prevalence of yaws infections and variation in exposure due to the timing of the penicillin campaigns. My results indicate that, following the penicillin campaigns, cohorts born in districts with higher initial yaws prevalence achieved higher education outcomes than prior generations when compared with cohorts from districts with lower yaws prevalence. The results are particularly robust for the female subsample, where I observe increases in educational attainment for cohorts born just prior to the penicillin campaigns.

In Chapter 4, I study the development of political partisanship, examining the plausibly random spread of the cocoa swollen shoot disease in the Gold Coast/Ghana in the 1940s. In 1948, the Watson Commission which investigated riots in colonial Ghana sparked by the cocoa swollen shoot pest noted the political motivations of the disturbances. In this chapter, I utilize novel data on cocoa farm acreages and the spatial variation in the spread of the swollen shoot virus to investigate the impact of the pest on the development of local political movements. Based on responses from the Afrobarometer surveys, I find that today, individuals in districts which historically experienced a high intensity of the disease pest report stronger anti-government opinions, and are more likely to attribute success in life to individual effort than government support. I trace the historical roots of these political views by examining electoral results from the 1956 Legislative Elections in colonial Ghana. Conditional on region fixed effects, and various pre-epidemic district controls, I observe that more adversely affected districts were more likely to vote against the new center-left (Nkrumahist) government. By 2000, with multiparty democracy, these areas still vote against the center-left (Nkrumahist) party. This partisan opposition has an impact on the allocation of resources today. Using an instrumental variable strategy, I examine the impact of government opposition on local government transfers received in various districts, with the historic intensity of the pest shock as an instrument. I examine possible violations to the exclusion restrictions of the 2SLS strategy by ruling out the impact of the cocoa swollen shoot disease on other economic and social outcomes. Based on the approach developed by Conley, Hansen and Rossi (2012), I also document that the 2SLS results remain robust to moderate forms of violations to the exclusion restriction assumptions.

Contents

Acknowledgments

I thank my family members in Accra and New York for their support. Their continued encouragement has been very useful during this time in graduate school.

I am grateful to my advisors, James Robinson, Nathan Nunn and Rick Hornbeck for their assistance and mentorship. My interests in studying the importance of institutions and historical development in Africa grew out of many useful conversations with my advisor, James Robinson. A devoted scholar, his breadth of knowledge in economics, politics, history, and anthropology, continues to inspire me. Nathan Nunn has been a helpful mentor throughout my graduate research work. He steered me towards interesting research questions at an early stage, and introduced me to many research networks. I am also grateful to Rick Hornbeck for his encouragement, for providing valuable feedback on my empirical work and for reviewing my draft papers in the course of these projects.

Many faculty members in Cambridge and elsewhere provided helpful advice along the way. At Harvard, I benefited from useful conversations with Robert Bates, Claudia Goldin, Asim Khwaja, Rohini Pande, David Yanagizawa-Drott, Richard J. Zeckhauser, and other members of the Harvard Development Lunch community. I am particularly grateful to Ryan Sheely at the Harvard Kennedy School for agreeing to review my dissertation. I also received financial support from Tavneet Suri and Chris Udry via the NSF/AERC/IGC grants for my research travels. During visits to Tanzania, I was hosted by faculty at the University of Dar es Salam — including Dr. Adolf Mkenda, Dr. Vincent Leyaro, Prof. Issa Shivji and Prof. Wangwe — who provided helpful practical and intellectual suggestions. I am also grateful to Dr. Agnes Kijazi, Faustine Tilya, Joseph Ndunguru and Jafari Chobo at the Tanzania Meteorological Agency (TMA) for providing access to Tanzania's rainfall data. Dr. Kofi Ahmed (University of Ghana) and Dr. Kingsley Asiedu (WHO, Geneva) assisted with access to the records of the yaws penicillin campaigns conducted by the Gold Coast/Ghana Medical Field Units.

In the course of my research, I benefited from interactions with several other students who helped in shaping my ideas and work along the way. The informal Harvard/MIT lunch group for students working on historical development, as well as the MIT Political Economy Breakfast seminar and the Harvard Development Lunch provided valuable peer support at various stages of my research work. I benefited from useful conversations with Marcella Alsan, Horacio Larreguy Árbesu, Jennifer Brea, Melissa Dell, James

Feigenbaum, Ameet Morjaria, Maksim Pinkovskiy, Tristan Reed, and Christopher Robert. Crayton Scott Walker (Harvard Map Library) and Giovanni Zambotti (Harvard Center for Geographic Analysis) introduced me to the use of GIS datasets.

Finally, I am grateful to James Habyarimana and Oyebola Olabisi for helpful advice during my research work. During visits to Nigeria, I benefited from useful conversations with Dr. Ngozi Okonjo-Iweala, Dr. Menachem Katz, Dr. Bright Okogu and Dr. Mansur Muhtar. In Boston, there was also a lively and entertaining group of friends at 73 Prince Street, including Evans Essienyi, Emmanuel Magani, James Habia, Thomas Kyei-Boateng, Margaret Mongare, Margaret Nyamumbo, Martin Mbaya, Edo Bedzra, David Sengeh, Atta Addo, and Kofi Amoako-Gyan.

Administrative and financial support in graduate school was made possible thanks to assistance from the Harvard GSAS Financial Aid Office, and the doctoral program directors, Louisa van Baalen and Nicole Tateosian. I am grateful for all their assistance.

Chapter 1

Introduction

The three essays in this dissertation are broadly concerned with understanding the importance of historical events and development programs on economic and social outcomes today. This introductory chapter briefly reviews the existing research literature in economic history and development, and summarizes the main findings of the three essays presented in this dissertation.

There has been renewed interest in understanding the historical determinants of long-run economic development (Nunn, 2009). Recent empirical research has highlighted the importance of factor endowments as well as the legacy of various colonial institutions in explaining comparative development today (Engerman and Sokoloff, 1997, 2002; Acemoglu et al, 2001, 2002; La Porta et al., 1997, 1998). Two strands of work may be identified in the recent research literature in economic history and development. The first group of work focuses on the importance of historical institutions and institutional persistence, while a second body of work examines specific channels, notably cultural norms, through which historical shocks have persistent impacts. The recent literature on the importance of institutions in explaining comparative development may be traced to the work of Acemoglu, Johnson and Robinson (2001, 2002) and La Porta et al, (1997, 1998), building on earlier contributions by North and Thomas (1973). Acemoglu et al (2001) exploit European settler mortality during the colonial period as an instrument for contemporary institutions, and show that these institutions are important in explaining cross-country income per capita today. Examples of institutional persistence in specific country contexts are provided by

Banerjee and Iyer (2005) and Dell (2010). Banerjee and Iyer (2005) examine the long-term impacts of land tenure institutions introduced in colonial India. They provide evidence of differences in the levels of agricultural investments and local public goods provision today, which may be traced to the type of colonial land tenure institution implemented. Dell (2010) also examines the legacy of the mining *mita* implemented in colonial Peru and Bolivia, and documents the long-term impacts of the *mita* in lowering household consumption today. An interesting component of this body of research examines the importance of state antiquity in explaining current economic outcomes (Bockstette et al. (2002)), and in the context of Africa, there is evidence of the importance of pre-colonial institutions for contemporary development outcomes (Gennaioli and Rainer, 2007; Michalopoulos and Papaioannou, 2011).

A related strand of research focuses on how changes in underlying cultural norms serve as a specific channel through which historical shocks can have persistent impacts today (Nunn, 2012). Cultural norms serve as heuristics to assist individual decision-making in uncertain settings. Recent empirical research provides examples of historical persistence operating via changes in cultural norms. For example, Nunn and Wantchekon (2011) document that ethnic groups in Africa which historically experienced high levels of the slave trade report lower levels of trust in others today based on data from the Afrobarometer Surveys. Alesina et al (2011) examine the origins of gender roles in the household. They examine the well-known Boserup hypothesis, and provide evidence that variation in farming technologies used in traditional societies resulted in male specialization in agriculture and female specialization in domestic activities, and are reflected in gender norms today. Guiso et al. (2008) also investigate social capital of Italian city states dating from 1000-1300 AD, and its persistent impacts on city-level measures of social capital today. Tabellini (2008) examines the impact of trust in others and belief in individual effort on comparative economic development in Europe today.

For economists and historians working on Africa, the recent empirical economics research work has revived previous scholarly debates on the determinants of long-run development in Africa (see Hopkins, 2009; Fenske, 2010). Despite methodological diferences, the increased availability of cartography-based datasets provides interesting research questions for both economists and historians, and enables more careful analyses of causal mechanisms relating to Africa's long-run economic development. This provides an opportunity to provide quantitative assessments of economic history in Africa, building on earlier contributions in the history literature by scholars such as Hopkins (1973), Austen (1987), Feinstein (2005) and Zeleza (1993). In particular, the study of the interaction of historical shocks with institutions and cultural norms raises new and interesting research questions in understanding the history of economic development in Africa.

The three essays in this dissertation aim at contributing to work on economic history and development in Africa. In the following three chapters, I successively examine the villagization experiment in Tanzania in the 1970s, disease eradication in the Gold Coast/Ghana in the 1950s using penicillin, and the impacts of the cocoa swollen shoot disease (in the 1940s) on political developments in the Gold Coast/Ghana.

In Chapter 2, I examine the subject of villagization in Tanzania, a major episode of development planning in post-independence Tanzania. A large literature in politics, anthropology and economics examined the subject of village formation in mainland Tanzania following its national independence in 1961 (Coulson, 1983; Hyden, 1980; McHenry, 1979; von Freyhold, 1979; Putterman, 1986; Collier et al., 1986). In the first chapter, I revisit this episode of Tanzania's economic history, focusing on the experience of developmental villages (*vijiji uya maendeleo*) introduced in mainland Tanzania over the period 1974-1982. Combining historic data on Tanzania from the 1970s with data from population censuses and recent national household surveys, I investigate whether variation in the intensity of the government's villagization program explains within-region variation in development outcomes today. I document that, in the short-run, developmental villages led to an increase in various educational outcomes, such as primary school completion rates, literacy rates, and total years of schooling. Today, districts which experienced a high share of developmental villages have greater availability of some public goods and report higher rates of participation in community activities, but citizens report a worse perception of corruption among government officials and greater rejection of one-party rule. Per capita household consumption is also significantly *lower* in districts with historically high levels of the treatment measure. To address potential endogeneity in village formation, I report instrumental variable results based on variation in ethnolinguistic fragmentation and the occurrence of droughts in the 1970s which facilitated the resettlement of peasants into villages. I conclude by providing some preliminary evidence on the lack of economic diversification as well as political alignment to the TANU/CCM party as possible channels for persistence.

In Chapter 3, I turn to the subject of disease eradication, and examine the impact of the successful control of a highly infectious tropical disease, yaws, in Ghana over the period 1956-1963. The availability of cheap, mass-produced penicillin following World War II, resulted in a mass treatment campaign by WHO/UNICEF aimed at controlling the prevalence of yaws and other bacterial infections. I examine the effect of this treatment campaign in which over 70 percent of the estimated Ghanaian population received a single dose of an intramuscular penicillin injection. Data collected by the WHO/UNICEF program before and after the campaign indicates that penicillin-based treatment resulted in an immediate reduction in the prevalence of infectious yaws among the Ghanaian population. Using a microsample from the 2000 Ghanaian census, I estimate a difference-in-difference model exploiting spatial variation in pre-treatment prevalence of yaws infections and variation in exposure due to the timing of the penicillin campaigns. My results indicate that, following the penicillin campaigns, cohorts born in districts with higher initial yaws prevalence achieved higher education outcomes than prior generations when compared with cohorts from districts with lower yaws prevalence. The results are particularly robust for the female subsample, where I observe increases in educational attainment for cohorts born just prior to the penicillin campaigns.

In Chapter 4, I study the development of political partisanship, examining the plausibly random spread of the cocoa swollen shoot disease in the Gold Coast/Ghana in the 1940s. In 1948, the Watson Commission which investigated riots in colonial Ghana sparked by the cocoa swollen shoot pest noted the political motivations of the disturbances. In this chapter, I utilize novel data on cocoa farm acreages and the spatial variation in the spread of the swollen shoot virus to investigate the impact of the pest on the development of local political movements. Based on responses from the Afrobarometer surveys, I find that today, individuals in districts which historically experienced a high intensity of the disease pest report stronger anti-government opinions, and are more likely to attribute success in life to individual effort than government support. I trace the historical roots of these political views by examining electoral results from the 1956 Legislative Elections in colonial Ghana. Conditional on region fixed effects, and various pre-epidemic district controls, I observe that more adversely affected districts were more likely to vote against the new center-left (Nkrumahist) government. By 2000, with multiparty democracy, these areas still vote against the center-left (Nkrumahist) party. This partisan opposition has an impact on the allocation of resources. Using an instrumental variable strategy, I examine the impact of government opposition on local government transfers received in various districts, with the historic intensity of the pest shock as an instrument. I examine possible violations to the exclusion restrictions of the 2SLS strategy by ruling out the impact of the cocoa swollen shoot disease on other economic and social outcomes. Based on the approach developed by Conley, Hansen and Rossi (2012), I also document that the 2SLS results remain robust to moderate forms of violations to the exclusion restriction assumptions.

Chapter 2

The Legacy of State Planning: Evidence from Villagization in Tanzania

2.1 Introduction

In *States and Power in Africa*, Jeffrey Herbst (2000) observes a recurring challenge concerning state development in Africa: that central authorities have often faced the difficulty of broadcasting their power throughout their territories. Since precolonial times, he argues, state development in Africa has been limited partly because of an unfavorable political geography characterized by low population density. Given the relative land abundance in many parts of Africa, scattered peasant populations presented a problem of state consolidation in newly independent African states in the 1960s. The ability of the state to raise taxes, to defend its territorial boundaries, and to provide public goods (enforcing property rights and providing social services) was hindered by large peasant populations in the hinterland that often tended to avoid the control of central governments.[1]

Since Chayanov (1966), the unique features of peasants in pre-capitalist societies has

[1]The existence of scattered peasant settlements in pre-colonial Africa may partly be the result of the dispersion of human populations as a means of escaping coercion (e.g. labour exploitation or taxation) by a centralized authority (see for example, McIntosh (1999, pp 22)). In the case of Tanganyika in the early twentieth century, many people moved away from formal settlements as a means of avoiding tribal wars and also escaping coercive enlistment by German officials for head porterage services during World War I (McHenry, 1979, pp 14; Iliffe, 1979, Chapter 8).

been noted in the literature, particularly the primary concern with subsistence production and catering to needs of the household. Peasant households in Africa were also different from smallholder farming communities in South Asia, where land scarcity and the need for investments in irrigation often required some dependence on the state or markets for survival. In contrast, peasant communities in Africa relied predominantly on rain-fed agriculture, and so remained independent of (”uncaptured by”) the state or other social classes such as capitalist landlords (Hyden, 1980, pp.31). The exercise in state formation in newly independent Africa necessarily involved bringing many such rural communities effectively under state administration.

This paper examines the effects of an historical experiment in state planning in rural Africa. Specifically, I examine the case of villagization which commenced in mainland Tanzania in 1973, and resulted in an experiment of developmental villages between 1973 and 1982 in rural parts of Tanzania. Villagization refers simply to the agglomeration of rural living units to facilitate state administration. However, these government-planned villages in Tanzania in the 1970s (termed as *developmental* or *registered villages*) involved more than the concentration of rural populations. These villages introduced state capacity at the local level by introducing village councils responsible for taxation, enforcement of property rights, and provision of public goods. The experiment of development villages was largely halted in 1982 following the repeal of the villages legislation and the commencement of an IMF economic liberalization program. State-led development planning and big-push theories served as the economic orthodoxy in many post-independence states in Africa. Therefore, the case of Tanzania, and its subsequent economic collapse, provides a poignant example of the legacy of state planning.

Using historical data from the 1978 Tanzania Census, I construct a measure of villagization as the fraction of a district’s population living in registered or developmental villages. I then investigate impacts of the villagization treatment on various short- and long-run outcomes. Using the IPUMS microsample for the 1988 Tanzania census, I first examine the short-run impact of the program on completion of primary education. I exploit variation in the intensity of villagization across Tanzanian districts and differences in exposure to villagization across birth cohorts due to the timing of the program, and document an increase in primary school completion for cohorts exposed to the program. Based on responses from the Tanzania Afrobarometer survey, I provide some evidence that a higher level of the treatment variable is associated today with greater political participation and greater support for democracy, but a rejection of one-party rule and worse perception of corruption among government officials.

For contemporary outcomes, I examine impacts of the program on household consumption, public goods provision and community participation. Using the 2009 Tanzania National Panel Household Survey, I find evidence of higher provision of some public goods and greater participation in community activities in treated districts today. However, there are significant consumption losses, and I find that increasing the district treatment intensity variable from zero to 1 results in approximately a 38 percent loss in per capita househod consumption today. The OLS regression results are conditional on region fixed effects, and remain robust to controlling for various pre-treatment covariates such as geographic characteristics (latitude, slope, altitude, mean and standard deviation of long-run precipitation), agricultural characteristics, health and education infrastructure, and per capita local government revenues.

A primary concern in my analysis is the problem of selection into the villagization treatment and the need to explain the source of the observed variation in the treatment measure. The historical account of Tanzania during this period highlight the occurrence of a drought in the mid-1970s which supported the government's villagization program (Hyden, 1980; Nyerere, 1977). The droughts in 1973-1975 facilitated peasant resettlement since with a bad harvest and the need for famine relief many peasants agreed to relocate to new government-planned villages. In addition, ethnographic accounts suggest that higher ethnolinguistic fragmentation in a district hindered cooperation and thus inhibited village formation (McHenry, 1979). I construct an instrument based on the interaction of the drought severity and pre-treatment ethnic fragmentation which helps to explain the observed variation in villagization treatment across districts. I conduct falsification tests with this instrument by showing that the interaction of ethnic fragmentation with similar rainfall shocks in the three-year period preceding, and immediately following, the period of villagization do not explain the observed variation in intensity of villagization. My instrumental variable estimates support the OLS results above.

Next, I explore the possible channels of persistence. To explain the severe consumption losses reported, I argue that economic diversification (i.e. a transition away from peasant agriculture) has been limited in areas which historically experienced high levels of villagization. This is supported by quantitative evidence based on the 2002 Tanzanian census. I also present some preliminary evidence that the treatment effect results in lower oppostion vote shares, suggesting that districts which experienced high levels of villagization in the 1970s remain loyal supporters of the ruling TANU/CCM party. This may partly explain the higher availability of public goods in these districts today.

This study contributes to several strands of research in the literature. First, it contributes to the broad research project examining the persistent effects of historical institutions (Acemoglu et al, 2001, 2002; Banerjee and Iyer, 2005; Dell, 2008; see Nunn 2009, for recent review), and the importance of past institutions in explaining comparative economic development in Africa (Gennaioli and Rainer, 2007; Huillery, 2009; Michalopoulos and Papaioannou, 2011). A related objective of this paper is to examine the impact of state-led planning in modifying cultural norms and individual perceptions on economic and political attitudes. In this regard, this paper contributes to the current research agenda examining how changes in cultural norms serve as a channel through which historical shocks have persistent effects (see Nunn, 2012 for a review). In a related vein, Alesina and Fuchs-Schündeln (2007) examine the impacts of German separation and reunification on preferences for redistribution in West and East Germany, while Jacob and Tyrell (2010) trace the erosion of social capital in East Germany to the density of state surveillance under the German Democratic Republic (GDR). In additional examples related to social capital and community participation, Guiso et al. (2008) also observed the persistence of social capital of Italian city states dating from 1000-1300 AD, and its impacts on city-level measures of social capital today. Tabellini (2008) also examined the impact of trust in others and belief in individual effort on comparative economic development in Europe today, using nineteenth century literacy rates and historical political institutions as instruments. Finally, Nunn and Wantchekon (2011) also document that ethnic groups in Africa which historically experienced high levels of the slave trade report lower levels of trust in others today based on data from the Afrobarometer Surveys.

Second, this project revisits an earlier body of social science research in economics and the multidisciplinary field of *Peasant Studies*, particularly examining village formation and development planning in Tanzania. This large body of scholarship had important contributions in the economics literature (Putterman, 1981, 1986; Collier et al, 1986; Ellis, 1982). Based on extensive data collection in Tanzania, Putterman (1981, 1986) provides a theoretical and empirical discussion on the design of an incentive structure for collective agriculture. Collier et al. (1986) surveyed households in eight regions in Tanzania, and examine patterns of rural inequality and the provision of social infrastructure in villages. Ellis (1982) documents the impacts of declining real prices and incomes in the agricultural sector on living stanards of peasant households. More recently, Putterman (1995a) discusses the effects of agricultural market liberalization on smallholder producers, as well as social capital and development in rural Tanzanian (1995b). Major debates in the peasant studies literature surrounded the extent of within-village rural inequality among the peasantry (Shivji, 1976; Ergas, 1980; von Freyhold, 1979; Raikes, 1978), on bureaucratic control (Raikes, 1975; Fortmann, 1980), on changes in land holdings following villagization (de Vries and Fortmann, 1979), and on the collapse of collectivised agriculture (Lofchie, 1978; Samoff, 1981). McHenry (1979), Hyden (1980), and Coulson (1982) provide useful accounts of this period of Tanzania's history. A useful bibliography of parts of this earlier literature is summarized by McHenry (1981).[2]

Third, this paper also provides lessons on the legacy of large-scale development planning projects which were widely implemented in the mid-twentieth century in many African states (Killick, 1976, 1983). The intellectual basis for national planning programs may be traced to structuralist development models following Rosenstein-Rodan's big-push approach (see Chenery (1971, 1975)). Development plans were generally aimed at tackling coordination failures in investments, and addressing differences in private and social valuations of various goods. The experience of development planning in Tanzania is reviewed by Rweyemamu and Mwansasu (1974, eds) and van Ardakie (1972). More recently, Scott (1999) examines the case of villagization in Tanzania, and provides a cautionary note on

[2]This Chapter is not intended to provide a normative assessment of Tanzania's villagization experiment which resulted in significant welfare losses and the country's economic collapse in the early 1980s. Rather, the goal is to study the legacy of this major episode of state planning, and provide an initial assessment of its impacts on various economic and social outcomes today. Data limitations also prevent me from examining the effects of villagization on land holdings in this present chapter, but this subject has been recently examined in the report of the Shivji Commission (see Shivji et al, 1994; Coldham, 1995).

the consequences of state planning.

The rest of this chapter is organized as follows. In section 2.2, I present a historical overview of post-colonial Tanzania, highlighting attempts by the government to encourage village formation and communal production. In section 2.3, I describe the various data sources. Next, I examine the impacts of the villagization program on educational outcomes and on political attitudes in sections 2.4 and 2.5 respectively. In section 2.6, I present results on household consumption, public goods provision and community participation, reporting both OLS and 2SLS estimates. I briefly discuss channels of persistence in section 2.7, and conclude in section 2.8.

2.2 Historical Overview of Village Formation

2.2.1 Phase I: 1961-1973

At independence in 1961, about 90 percent of the mainland Tanzanian population lived in the rural countryside in scattered hamlets which were not organized into formal villages. A visiting World Bank team in 1959 urged the country's leadership to consider the settlement of sparsely populated areas which had agricultural development potential (World Bank, 1961).[3] The country's low population density had often been highlighted in official records by German and British colonial staff as an obstacle to economic development of the colony. Julius Nyerere, president of the newly independent Tanzania, however advocated a more voluntary resettlement plan in which exhortations and inducements (such as provision of clean water and schools) served as the major tools to encourage village formation. The use of exhortation to promote villagization and thus obtain economies of scale in the provision of social infrastructure and agricultural machinery was clearly evident in Nyerere's inaugural address:

> So if you ask me what our Government is planning to do during the next few years the answer is simple. For the next few years Government will be

[3] Other proposals were also offered to the Tanganyika leadership, for example, one based on the Israeli model of *moshav* village communities which were similar to the *kibbutzim* but involved a lesser degree of communalization (Kaplan, 1961).

> doing all it can to enable the farmers of Tanganyika to come together in village communities. And if you ask me why the Government wants us to live in villages, the answer is just as simple: unless we do we shall not be able to provide ourselves with the things we need to develop our land and to raise our standard of living. We shall not be able to use tractors; we shall not be able to provide schools for our children; we shall not be able to build hospitals, or have clean drinking water, it will be quite impossible to start small village industries, and instead we shall have to go on depending on the towns for all our requirements; and even if we had a plentiful supply of electric power we should never be able to connect it up to each isolated homestead. (*President's Inaugural Address*, Nyerere (1966) pp 183-4).

There were isolated settlement schemes which autonomously formed in some parts of the country, often with a nucleus of youth activists of the governing TANU party. The approach to rural development in the period following independence involved largely the use of *persistent persuasion* to encourage progressive (yeomen) farmers to adopt modern agricultural techniques in cultivating the main cash crops (sisal, coffee, cotton, tobacco and maize). There were also a few government-planned settlements which were developed in mainland Tanzania (Hyden, 1980). With the exception of the Ruvuma Development Association (RDA), many such village settlements failed. In a landmark speech in 1967, called the *Arusha Declaration*, Nyerere outlined a socialist vision for the Tanzanian state, which involved a policy of national economic self-reliance, nationalization of commercial farms and industries, and the promotion of socialism in the villages (*ujamaa vijijini*).

The idealized ujamaa villages would involve a collection of peasant households working on communal farms organized using modern agricultural methods, and generating surplus income to finance various social infrastructure. However, as the ujamaa philosophy was seen as extension of traditional norms of cooperation in villages, Nyerere's TANU party prevented the use of compulsion in starting new villages. In the years following the Arusha Declaration, various inducements, such as the provision of social infrastructure (schools, clean water and dispensaries), were used to promote village formation. By 1973, official estimates suggest that nearly 10 percent of rural households lived in some form of a village.

2.2.2 Phase II: 1973-1982

However, given the slow pace of village formation, in November 1973, the Nyerere administration announced the mandatory resettlement of all peasant households into villages by the end of 1976. Villages were to be established within the stipulated 35-month window, and were to be termed as registered or developmental villages (*vijiji uya maendeleo*). The major relocations were broadly termed as *Operation Sogeza* (meaning moving in Swahili). In a span of 20 months, there was a remarkable transformation of the Tanzanian countryside, with nearly 85 percent of all rural households now living in some form a village. New villages were rapidly planned and with a minimum of 250 families needed. The actual implementation of *Operation Sogeza* varied across regions, but in most cases operations were divided into a number of stages with a time-table involving: identifying existing location of farmers, selecting new village sites, and final resetttlement (McHenry, 1979).

What factors accounted for the relative success of the villagization resettlements across districts? While there are potentially many factors which contributed to the observed intensity of villagization in each district, I focus here on two main factors highlighted as important in the historical literature: the use of compulsion by local TANU party cadres, and the occurrence of widespread droughts in Tanzania during the mid-1970s. First, there are several documented cases in which local TANU party activists used compulsion in moving local residents into new villages (Coulson, 1982:250-253). To account for the relative effectiveness of local TANU party activists in my emprical analyses, I specifically control for the share of votes received by the TANU Party in the 1970 Tanzania National Presidential Elections. Second, drought conditions which hit several parts of Tanzania in the mid-1970s appear to have been important in causing many peasant households to relocate to planned villages. The anthropological literature has often highlighted the independence of the peasant mode of production in Tanzania, as many rural households often engaged in subsistence rain-fed agriculture without much concern for the broader activities of the nation-state. The droughts of the mid-1970s however created the need for famine relief which was provided in newly planned villages, and for farmers without a harvest, the costs of relocation into villages was lowered. I revisit the use of the drought severity as an instrument for the intensity of villagization at the district level in Section 5 on my instrumental variable 2SLS results.

2.2.3 Institutional Features of Developmental Villages, 1973-1982

The new developmental villages were not simply agglomerations of rural households but involved institutional changes legislated in the *Villages and Ujamaa Villages Act* of July 1975 (Verhagen, 1980). The reforms in village administration in rural Tanzania approximately spanned the period 1974-1982 until the repeal of the Act, the commencement of an economic liberalization program, and the announcement of President Nyerere's imminent resignation (Shivji and Peters, 2003; Nord et al., 2009). Chieftaincies which had served as a focal point of native administration during the colonial period had been abolished by the TANU administration in 1963, and all lands nationalised to be under the control of the Presidency[4]. Villagization therefore reorganized the Tanzanian countryside with four main institutional changes (Verhagen, 1980; Fortmann, 1980; Collier, Radwan and Wangwe, 1986).

First, all adult residents in a village became members of a Village Assembly, which met at least once a year. Policy-making at the village level was delegated to a Village Council (with a Chairman and CEO), and elected by the Village Assembly. Further, the Village Council established five committees responsible for finance, production, education, works and defense. Second, once registered, the developmental village was legally entitled to contract loans and engage in economic activity such as marketing of crop produce. Co-operatives which had previously dominated the crop marketing chain were abolished, with their functions now handled by village councils. Third, the village also mobilized the local community to support the provision and maintenance of social infrastructure. With all lands effectively nationalised, the Village Council was also empowered to allocate land among private cultivators. Finally, the village became the center of development planning. Village development plans were designed and forwarded to the district and regional administrators for review and implementation. In 1978, about 4000 development managers were appointed and deployed to live in registered villages to assist in preparation of village

[4]Chiefaincies were abolished following the legislation of the African Chiefs Ordinance (Repeal) Act, 1963.

development plans.

This episode of village participatory development was largely halted in the early 1980s following Tanzania's economic collapse. Village councils lost their significance as units for economic planning or for governance and decision-making. Specifically, in 1982, the Villages and Ujamaa Village Act was repealed and replaced with the more orthodox Local Government (District Authorities) Act (1982, No 7). Decision making in these new local governments was directed from central government officials, and suffered from the "lack of democratic participation from below" (Shivji and Peters, 2003, pp. 13). Nyerere also announced his decision not to contest the 1985 presidential election, while Tanzania embarked on an economic liberalization program supported by the IMF (Nord et al., 2009; Freund, 1981). The IMF adjustment program involving liberalization of agricultural markets, removal of domestic price controls, and reform of parastatals withdrew many government subsidies previously channeled to village councils (Skarstein, 2005).

2.3 Data and Summary Statistics

2.3.1 Data

I examine the long-run effects of the late 1970s villagization by examining their impacts on education, consumption, community participation and public good provision in Tanzania today. The villagization treatment varies at the district level and is constructed using data from village gazetteers in the 1978 Tanzania National Census. I focus on the district, as this administrative unit remained the center of political and economic activity in the period following independence (Iliffe, 1975). Using 1978 boundaries, I calculate the treatment measure for each district as the share of the population living in developmental or registered villages:

$$T_d = \frac{P_{d,r}}{P_{d,t}} \tag{2.1}$$

where $P_{d,r}$ refers to population in all registered villages in district, d and $P_{d,t}$ is the total population in district, d. Figure 2.1 provides a graphical illustration of the spatial variation in the treatment intensities.

For outcome measures, I use data from a variety of sources: the IPUMS microsample for the 1988 and 2002 Tanzania National Census, the 2000 Tanzania Household Budget Survey, the 2008/2009 Tanzania National Household Panel Survey, and the Afrobarometer Surveys for Tanzania (Rounds 1, 3 and 4).

The IPUMS dataset provides a 10 percent microsample of the 1988 Tanzania census, with about 2.4 million persons. In the subsection examining short-run education outcomes, I restrict my analyses to cohorts aged between 1 and 24 in 1974 comprised of about 820,000 individuals in mainland Tanzania. I examine contemporary occupational outcomes using IPUMS microdata from the 2002 Tanzania census containing about 3.7 million observations. Next, consumption data is obtained from the Tanzania Household Budget Survey (THBS) conducted in 2000. Although there are more recent consumption surveys for Tanzania, the 2000 HBS provides data covering the largest available sample of 22,178 households. I focus on the total consumption per adult equivalent (standardized for a 28-day period, and stated in nominal Tanzanian Shillings) as my main consumption outcome. This is the standard consumption aggregate measure used in poverty analyses in Tanzania.

For outcomes on participation in community activities and public goods provision, I use the Tanzanian National Panel Survey (2008/9). The 2008/9 Tanzania NPS is the first wave of a nationally representative household survey conducted by the Tanzania National Bureau of Statistics, and forms part of the World Bank's Living Standard Measurement Surveys (LSMS). The survey includes separate units for a household survey, a community module, and an agricultural survey. The household survey is comprised of 3,280 households and 16,709 individuals. The household module records respondents' districts of birth, thus enabling me to link migrants to their initial districts of birth. I also examine outcome variables related to perceptions on corruption, democracy and national identity which are obtained from the Afrobarometer Surveys for Tanzania. The Afrobarometer Surveys are nationally representative surveys of voting age citizens in Africa, and have been previously used elsewhere in the literature (see, for example, Nunn and Wantchekon, 2009). I utilize results from Rounds 1, 3 and 4 of the Tanzania surveys which provide district identifiers for respondents.

In examining electoral outcomes today as a possible channel of persistence, I utilize

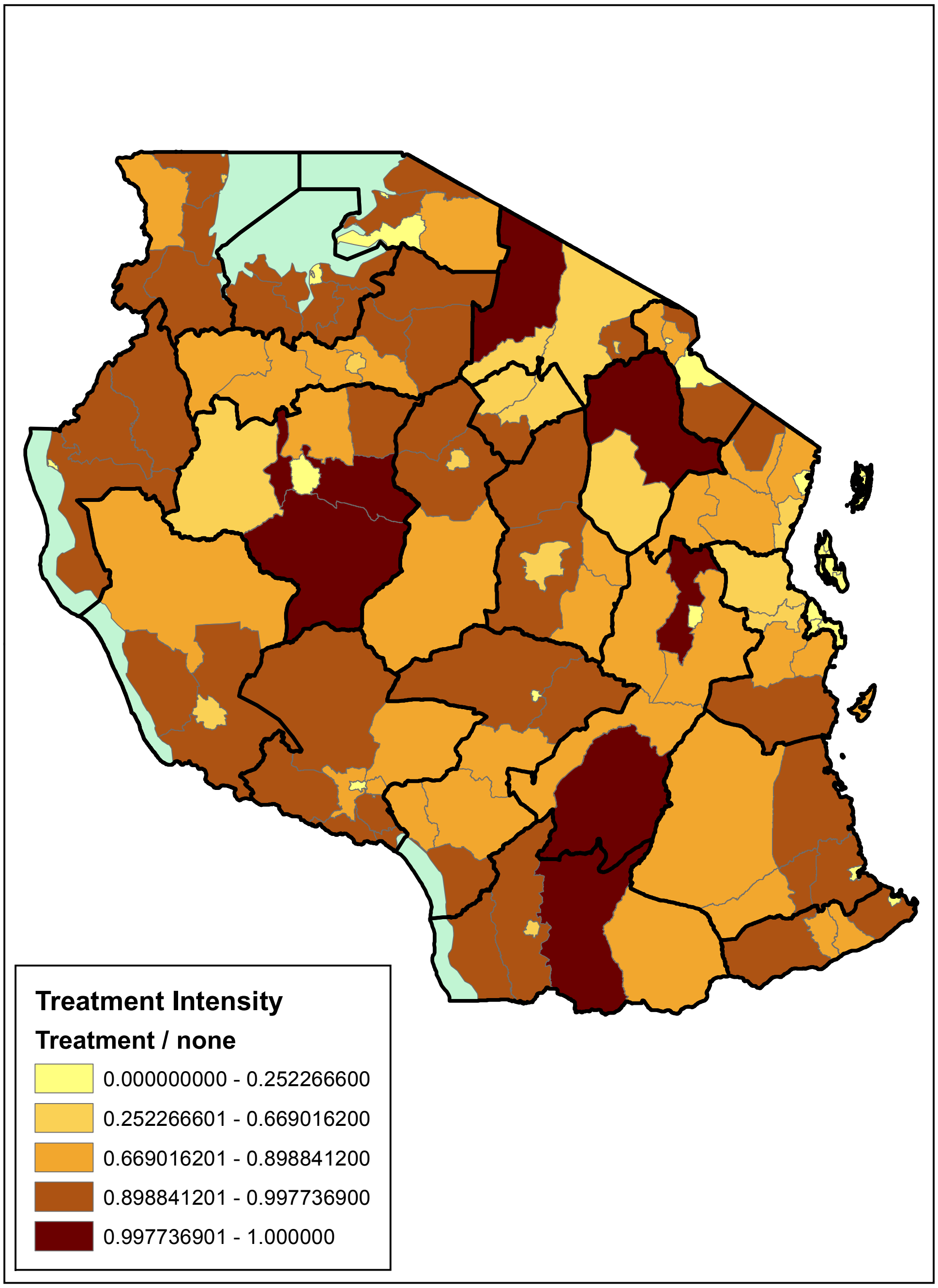

Figure 2.1: **Treatment Intensity Rates** (fraction of population living in registered villages). Region boundaries are darkened.

results of the 2010 Tanzania presidential election results obtained from the National Electoral Commission of Tanzania. I also include district-level election results from the 1970 Tanzanian National Election as a control for the popularity of Nyerere's TANU Party prior to village formation (see Election Study Committee, 1970). The 1970 electoral results provide district-level outcomes on voter turnout and the fraction of the population voting *yes* in support of Nyerere in the one-party election. Local government tax revenues (for agricultural produce cess) are also obtained from the Tanzanian Government PMO-RALG office.

Precipitation data is obtained from the Tanzania Meteorological Agency (TMA). Previous research on Africa with climate data often utilized global gridded datasets in which rainfall data is obtained by spatially interpolating information from selected rainfall stations. For most parts of Africa, the low density of stations used in the spatial interpolation implies that very local climate variability may not be adequately detected. For example, the East Anglia Climate Research Unit (CRU) data for Tanzania is based on observations for only three stations (Dar es Salaam, Songea and Tabora). Therefore, in this project, in order to investigate the local variation in drought conditions needed for my instrumental variables estimates, I obtained monthly precipitation data on 108 rainfall stations in Tanzania covering most of the period 1960-2010. Many rainfall stations in the TMA network have been in existence since the colonial period under the East Africa Meteorological Agency (see Johnson, 1962; Nieuwolt, 1977). I calculate district-level mean precipitation for the main long-rains planting season (*masika*) which approximately covers March to June each year.

Finally, I also control for various district pre-treatment characteristics using data compiled by Jensen and Mkama (1968). The authors served in the UN Team in Physical Planning in the Tanzanian Ministry of Lands, Settlement and Water Development. They compile district-level data based on various primary sources available in Tanzania in the 1960s (see Jensen and Mkama, 1968, pp 81). Specifically, they provide district-level data on demographic characteristics, health infrastructure (number of hospital beds per inhabitant), schools and education enrollment rates, local government revenues, agricultural characteristics and economic activity. Geographic controls are also used, including the long-run precipitation (mean and standard deviation), latitude, mean area weighted elevation and slope, distance from the centroid of each district to the coast and to the nearest railway line (in 1975).

Details of the data compilation process are provided in Appendix A.

2.3.2 Descriptive Statistics

Table 2.1 provides a summary of district characteristics. The mean level of the villagization treatment variable across all districts is 0.727. I group all districts into three categories of high, medium and low treatment areas and also report additional pre-treatment characteristics. Most pre-treatment characteristics are fairly balanced across all three categories. The notable exception is the level of ethnolinguistic fractionalization, ELF_d. Districts with high levels of village treatment were less ethnically diverse. High treatment districts also tended to have a slightly lower share of moslems in the population. Tanzania was very sparsely settled with mean population density of 32 persons per square mile (based on the 1967 population census), and with high treatment districts tending to have slightly higher density of persons prior to village formation. In subsequent regressions, I control directly for the effects of pre-treatment ethnolinguistic fragmentation and various pre-treatment characteristics. Support for the ruling TANU party was also uniformly high across all districts in 1970. About 66 percent of registered voters, and 97 percent of total votes cast in the 1970 Presidential Election voted in support of the ruling TANU party.

2.4 Short-run Effects on Education

I begin by studying the expansion of educational outcomes in Tanzania in the 1970s which was viewed as a central component of the villagization program (Nyerere,1977; Mbilinyi, 1976). In contrast to other large education expansion programs financed by resource booms during this period (e.g. in Indonesia (Duflo, 2001), in Nigeria (Osili and Long, 2008)), Tanzania's program was achieved via village councils in development villages using informal labor for the construction of primary schools and teachers' quarters under the direction of regional government officials (Stabler, 1979; Sheffield, 1979). The sudden implementation of the villagization program (1974-1976) and its variation across

TABLE 2.1: CHARACTERISTICS OF DISTRICTS

Variables	All Districts (1)	Low Districts (2)	Medium Districts (3)	High Districts (4)	*P*-value (2) vs (3) (5)	*P*-value (3) vs (4) (6)
Treatment	0.727	0.357	0.880	0.965	0.0000	0.0000
(Frac. of pop. in registered villages, 1978)						
Demographic Characteristics in 1967						
Share of Moslems	0.293	0.403	0.267	0.202	0.2445	0.1882
Share of Christians	0.324	0.307	0.312	0.352	0.6848	0.2088
Ethnolinguistic fractionalization	0.610	0.738	0.624	0.460	0.0196	0.0181
Population density (persons per sq. m)	32.42	30.09	30.87	36.53	0.9176	0.6401
Geographic Controls						
Latitude	-5.661	-5.609	-5.866	-5.506	0.6316	0.6566
Longitude	35.24	36.10	35.35	34.21	0.4522	0.0542
Altitude	1,005	848.0	1,017	1,160	0.4082	0.1284
Slope	3.639	3.559	3.446	3.925	0.5932	0.3861
Distance to railway	107.3	75.09	126.6	121.8	0.0736	0.7706
Distance to coast	510.7	392.8	499.5	648.2	0.3387	0.0401
Mean rainfall (total for *masika* season, in mm)	492.71	520.67	425.74	529.71	0.1359	0.2243
Std. deviation of rainfall (in mm)	156.73	154.97	145.64	169.72	0.6730	0.4840
Social infrastructure in 1967						
Inhabitants per dispensary	10.10	11.46	9.851	8.912	0.4267	0.3739
Hospital beds per capita	2.084	4.018	1.125	1.010	0.0819	0.6600
School enrollment rate	76.04	81.84	72.42	73.58	0.5418	0.9712
Agricultural Characteristics in 1967						
Mean land area	1.529	1.808	1.514	1.247	0.3521	0.7043
Fraction of land under smallholder farms	0.0954	0.0940	0.0915	0.101	0.9922	0.8510
Cattle per capita (e-03)	0.919	1.20	0.708	0.837	0.3435	0.8962
Sheep per capita (e-03)	0.277	0.461	0.173	0.188	0.1736	0.7925
Goats per capita (e-03)	0.389	0.554	0.309	0.295	0.2405	0.9017
Donkeys per capita (e-06)	15.7	26.9	13.4	6.04	0.2891	0.2836
Pigs per capita (e-06)	1.21	0.874	1.27	1.50	0.6242	0.7707
Local Government Revenues						
Local government revenue per capita	8.56	8.42	8.51	8.76	0.7687	0.9518
(in Tz. Shillings x e-03)						
National Election Results, 1970						
Registered to eligible voters ratio	0.744	0.738	0.747	0.746	0.4001	0.5664
Voter turnout (voters to eligible pop. ratio)	0.513	0.506	0.526	0.507	0.2402	0.2279
Share of votes for Nyerere among total votes cast	0.969	0.969	0.968	0.970	0.3361	0.4649
Share of votes for Nyerere among registered voters	0.663	0.653	0.669	0.668	0.4229	0.8348
Share of votes for Nyerere among eligible population	0.494	0.486	0.502	0.494	0.8230	0.6403
Observations	90	31	30	29		

NOTES: Data compiled from Tanzania National Census (1978); Jensen and Mkama (1968); the Election Study Committee (1974). Additional information available in data construction appendix.

districts induces differences in exposure across various year of birth cohorts. This motivates a difference-in-differences estimator of the treatment effect as in Duflo (2001). The program commenced in 1974, and Tanzanian government policy stipulated school enrollment for children aged 7-13 years. Therefore children aged 1-6 years in 1974 were fully exposed to the villagization experiment whereas cohorts aged 14-19 fall out of the prescribed education window and may have made their primary education decisions prior to the commencement of the program[5]. The IPUMS subsample for the 1988 Tanzania census provides educational attainment and district of residence on about 2.3 million observations. I assume that the district of current residence is highly correlated with the district of education.

An overview of the results of the program on completion of primary education is presented in the two-by-two matrix in Table 2.2. High and low treatment areas are respectively districts that fall in the top and bottom quintiles of the villagization distribution. Panel A presents the experiment of interest comparing primary school completion rates for children aged 1-6 years in 1974 (and thus were fully exposed to the villagization experiment during their primary school years), to those aged 14-19 years who had little or no exposure. Primary education rates increase over the period in both high and low treatment areas by 0.29 and 0.18 points respectively. Under the identifying assumption that high and low treatment areas would have experienced similar increases in educational attainment, the difference in differences may be interpreted as the causal effect of the program. As shown in Table 2.2, primary school completion rates increased by about 11.5 percentage points more in high treatment areas compared with low treatment areas during the villagization period.

In Panel B, I provide results for a control experiment which compares education completion for two cohorts of children, aged 14-19 and 20-25 in 1974. As both cohorts were not exposed to the program during their years of primary education, we would not expect to see any significant difference between primary completion rates for these two groups. The difference in difference estimate in Panel B is about 0.9 percentage points, and not significantly different from zero — suggesting that primary school completion would not

[5] Delayed enrollment and grade repetition implies that some of the older cohorts would be treated, and in this case should bias my results downwards against finding a treatment effect.

TABLE 2.2: MEANS OF EDUCATION BY COHORT AND LEVEL OF VILLAGIZATION PROGRAM

	Dummy for primary school completion		
	Low Treatment Areas	High Treatment Areas	Difference
Panel A: Experiment of Interest			
Aged 1 to 6 in 1974	0.756	0.694	
Aged 14 to 19 in 1974	0.579	0.403	
Difference	0.176	0.291	0.115 (0.005)
Panel B: Control Experiment			
Aged 14 to 19 in 1974	0.579	0.403	
Aged 20 to 25 in 1974	0.451	0.264	
Difference	0.129	0.138	0.009 (0.006)

NOTES: Data is obtained from IPUMS microsample of 1988 Tanzania Census. Villagization program commences in 1974 and primary school attendance spans ages 7-13 years. In the main experiment, I compare children aged 1-6 years in 1974 who are fully exposed to the program, those aged 14-19 years in 1974 who are not exposed to the program. The control experiment compares children aged 14-19 in 1974 to those aged 20-25 in 1974. Outcome variable is a dummy equal to 1 for primary school completion.

have evolved differently in high and low treatment areas in the absence of the program.

Below, I provide more precise estimates of the effect of the program in a regression framework. Consider the fixed effects regression:

$$P_{idc} = \alpha_d + \beta_c + (T_d * E_i)\gamma_1 + (X_d * E_i)\delta_1 + \epsilon_{idc} \tag{2.2}$$

where P_{idc} is a dummy equal to one for primary school completion, E_i is a "young" exposure dummy equal to one if the individual was 6 years or younger in 1974, and thus exposed to the program, α_d and β_c are respectively fixed effects for districts of birth and year of birth, T_d refers to the villagization intensity in district d, and X_d is a vector of controls of pre-treatment characteristics for district d. The identification assumption here requires that no omitted time-varying and district-specific factors are correlated with the treatment intensity of the program. Therefore, to address possible omitted variables, I present results controlling for the interaction of the "young" dummy with various pre-treatment district characteristics such as primary school enrollment rates, health infrastructure, demographic, agricultural and geographic characteristics.

Table 2.3 presents the results, comparing cohorts aged 1-6 years in 1974 to those aged 14-19 years in 1974. The baseline regression without controls is presented in column 1, while in columns (2), (3) and (4) I successively control for pre-treatment education enrollment, health characteristics and finally all controls. The results indicate that moving from zero to full villagization increased the probability of primary school completion by about 13 percentage points. In Panel B, I repeat this exercise but now comparing cohorts aged 14-19 years in 1974 to those aged 20-25 in 1974 (i.e. the control experiment in Table 2.2 above). In this case, the coefficients are are much smaller, and not significant.

Next, I extend the analysis in equation (2.2) above to examine impacts of the treatment on education completion for various age cohorts. I run the following regression:

$$P_{idc} = \alpha_d + \beta_c + \sum_{l=1}^{24}(T_d * b_{il})\gamma_{1l} + (X_d * E_i)\delta_1 + \epsilon_{idc} \tag{2.3}$$

where b_{il} refers to a year of birth dummy for individual i. The omitted dummy category is for l=25. For each cohort, the coefficient γ_{1l} provides estimates for the impact of the villagization treatment compared to the omitted category or control group (i.e. for l=25). We would therefore expect no impact of the program on cohorts aged 14 and above in 1974 who would have completed their primary education prior to the introduction of development villages. Similarly, we expect the effect of the treatment to be positive for cohorts aged 13 and below (in 1974) and for this effect to increase for younger cohorts. Figure 2.2 displays these coefficients and provides evidence in support of this observation (see Appendix Table A1). The coefficients on the interaction of age cohort and treatment intensity are positive and significant for younger cohorts up to cohorts aged 14 in 1974. For older cohorts, aged between 15 and 24 in 1974, the coefficient is smaller and no longer significant. The rise in the coefficients begin for cohorts aged 17, although the 95 percent confidence intervals include zero. As expected, the coefficient is positive and increases in magnitude, and becomes significant precisely for children aged 13 years and below in 1974. The positive coefficients for cohorts aged 14-17 may be the result of these cohorts being partly treated if some older individuals were still enrolled in school in 1974. This may be likely given previous evidence of delayed primary enrollment in Tanzania (Bommier and Lambert, 1999).

In Table 2.4, I present results examining additional educational outcomes. I repeat the regression in equation (2.2) but now with literacy (defined as ability to read *and* write in any language) and total years of schooling as outcome variables. Consider the results reported in column (1) of Table 2.4, Panel A. The first row repeats results for the probability of primary school completion in Table 2.3. I find that literacy increases by 8.85 percentage points, while years of schooling completed increase by approximately 1.1 years. The results in increased educational outcomes also remain when I consider male and female subgroups in Panels B and C of Table 2.4. Overall, the discussion in this section provides some evidence that the villagization program yielded some short-term impacts in educational attainment.

TABLE 2.3: EFFECT OF PROGRAM ON PRIMARY EDUCATION COMPLETION

Dependent variable: Dummy for Completion of Primary Education	(1)	(2)	(3)	(4)
Panel A: Experiment of Interest (cohorts aged 1-6 or 14-19)				
Young cohorts are aged 1-6				
*Young * TreatmentIntensity*	0.130***	0.129***	0.118***	0.0832***
	(0.0259)	(0.0259)	(0.0276)	(0.0198)
Observations	469,034	469,034	459,164	446,484
R-squared	0.119	0.119	0.120	0.122
Panel B: Control Experiment (cohorts aged 14-19 or 20-25)				
Young cohorts are aged 14-19				
*Young * TreatmentIntensity*	0.0154	0.0162	0.0163	0.00853
	(0.0117)	(0.0117)	(0.0132)	(0.0138)
Observations	293,758	293,758	291,599	283,549
R-squared	0.089	0.089	0.090	0.089
Young dummy * primary enrollment rate in 1970	NO	YES	YES	YES
Young dummy * health infrastructure in 1970	NO	NO	YES	YES
Young dummy * other pre-treatment district characterstics	NO	NO	NO	YES

*** $p<0.01$, ** $p<0.05$, * $p<0.1$. Dependent variable is dummy for completion of primary education. Data is obtained from IPUMS microsample of 1988 Tanzania Census. All columns include fixed effects for year of birth cohorts and districts. Controls enter the regression interacted with the young dummy. Column 2 controls for pre-treatment primary enrollment in 1970. Column 3 includes controls for health infrastructure in 1970. Column 4 includes various geographic and demographic controls (for latitude, slope, altitude, ethnic fragmentation, share of christians, livestock populations, and local government revenues - see Jensen and Mkama, 1968). Robust standard errors are in parentheses and clustered at the 1978 district boundaries.

TABLE 2.4: EFFECT OF PROGRAM ON EDUCATIONAL OUTCOMES (CONTD.)

	(1)	(2)	(3)	(4)
PANEL A: FULL SAMPLE				
Primary education completion	0.130***	0.129***	0.118***	0.0832***
	(0.0259)	(0.0259)	(0.0276)	(0.0198)
Literacy	0.0885***	0.0858***	0.0849***	0.0588***
	(0.0187)	(0.0194)	(0.0221)	(0.0164)
Years of schooling	1.079***	1.067***	1.025***	0.826***
	(0.198)	(0.201)	(0.221)	(0.167)
PANEL B: MALE SUBSAMPLE				
Primary education completion	0.131***	0.130***	0.117***	0.0899***
	(0.0237)	(0.0244)	(0.0260)	(0.0247)
Literacy	0.0584***	0.0566***	0.0573***	0.0310**
	(0.0135)	(0.0144)	(0.0174)	(0.0146)
Years of schooling	1.074***	1.064***	1.045***	0.903***
	(0.173)	(0.179)	(0.206)	(0.196)
PANEL C: FEMALE SUBSAMPLE				
Primary education completion	0.103***	0.102***	0.0937***	0.0632***
	(0.0284)	(0.0282)	(0.0309)	(0.0233)
Literacy	0.0897***	0.0856***	0.0843***	0.0682***
	(0.0268)	(0.0270)	(0.0304)	(0.0241)
Years of schooling	0.864***	0.848***	0.797***	0.633***
	(0.227)	(0.229)	(0.252)	(0.199)
Young dummy * primary enrollment rate in 1970	NO	YES	YES	YES
Young dummy * health infrastructure in 1970	NO	NO	YES	YES
Young dummy * other pre-treatment district characterstics	NO	NO	NO	YES

*** p<0.01, ** p<0.05, * p<0.1. Table reports coefficients on *Young * TreatmentIntensity*. Data is obtained from IPUMS microsample of 1988 Tanzania Census. All columns include fixed effects for year of birth cohorts and districts. Controls enter the regression interacted with the young dummy. Column 2 controls for pre-treatment primary enrollment in 1970. Column 3 includes controls for health infrastructure in 1970. Column 4 includes various geographic and demographic controls (for latitude, slope, altitude, ethnic fragmentation, share of christians, livestock populations, and local government revenues - see Jensen and Mkama, 1968). Robust standard errors are in parentheses and clustered at the 1978 district boundaries.

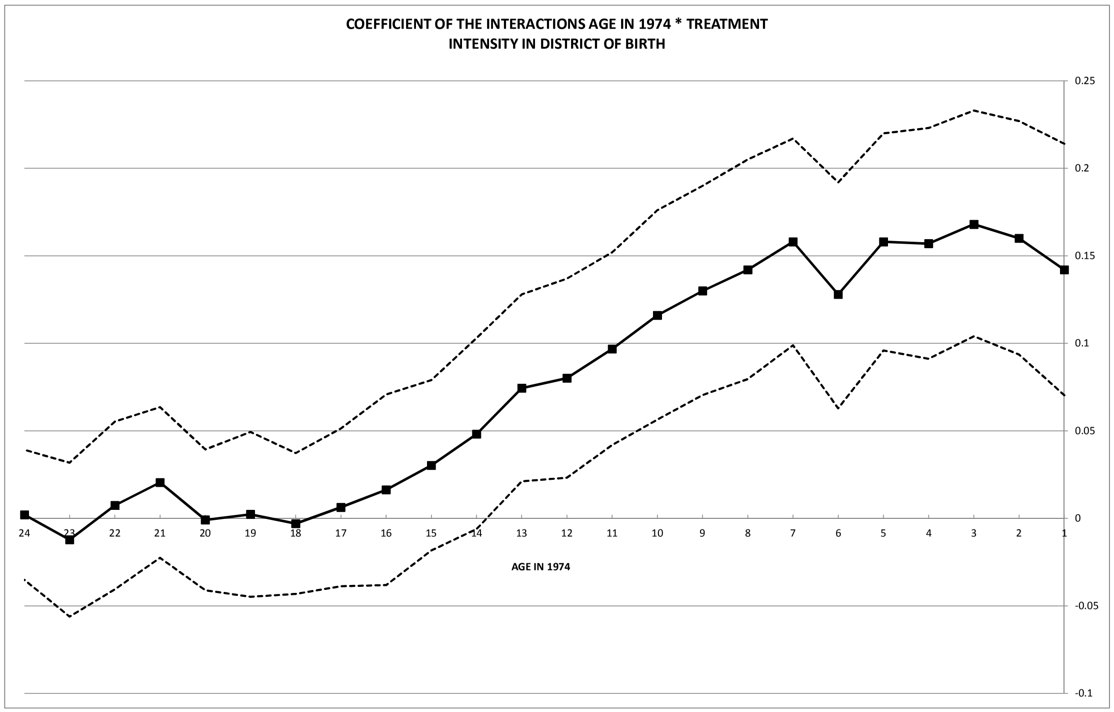

Figure 2.2: Increases in Primary School Completion Rates. This chart plots the coefficients γ_{1l} of $(T_d \times b_{il})$ in Equation 2.3. Each point provides an estimate of the impact of the villagization treatment on primary education completion rates for the relevant age cohort.

2.5 Effects on Political Attitudes

Political socialization was a secondary objective of the villagization campaigns, and aimed at developing norms of citizenship in the new nation (Nyerere, 1977). Nation-building in newly independent African states often required programs aimed at re-orienting citizens away from parochial ethnic loyalties (Koff and von der Muhll, 1967). In the case of Tanzania, specific policies of nation-building implemented by the post-independence government included emphasizing civic education in the public school curriculum and promoting Swahili as a national language (Prewitt et al, 1970, 1971; Miguel, 2004). In this section, I turn to recent Afrobarometer Surveys, and examine changes in political attitudes of cohorts born before and after the villagization campaigns. Specifically, I examine outcomes related to political participation, support for one-party rule, support for democracy, perceptions of corruption among government officials and perceptions of ethnic versus national identity. The Afrobarometer surveys are conducted for citizens aged 18 years and above. In this analysis, I omit individuals aged between 38 and 48 in 2008 (i.e. born in the 1970s during the villagization campaigns) and compare older cohorts, aged 48 and above, to younger cohorts aged 38 and below.

I run a regression similar to equation (2.2) above:

$$Y_{idc} = \alpha_d + \beta_c + (T_d * E_i)\gamma_1 + (X_d * E_i)\delta_1 + \epsilon_{idc} \tag{2.4}$$

where Y_{idc} is the outcome variable of interest; E_i is a "young" if the individual was aged between 18-38 in 2008, and thus born following the villagization programs, α_d and β_c are respectively fixed effects for districts of birth and year of birth, T_d refers to the villagization intensity in district d, and X_d is a vector of controls of pre-treatment characteristics for district d. The coefficient, γ_1 on $(T_d * E_i)$ therefore provides the difference-in-difference estimate comparing the mean outcome variable for older cohorts aged 48-78 years to a younger cohort aged 18-38 years, in high versus low treatment districts.[6]

[6] In additional results reported in Appendix C, Tables C1-4, I extend the analyses by categorizing the old and young cohorts into further age bins, and then running a regression similar to equation (2.4) above. In this case, the dummy E_i takes on four values for cohorts born between 1980-1989, 1970-1979, 1960-1969 and 1950-1959. The omitted category is for cohorts born prior to 1950.

The results on political attitudes are reported in Tables 2.5-2.8. All outcome variables are coded as dummies for whether the respondent agrees or disagrees with the question posed. For each table, in moving from columns (1) to (6), I successively introduce controls for geographic characteristics, and pre-1970 demographic composition, agricultural characteristics, education enrollment and local government revenues. The difference-in-difference estimates in Table 2.5 indicate that younger cohorts in more treated districts, are more likely to have contacted their local government or civil service officials, and more likely to be able to correctly name their local MPs.

In Table 2.6, I find that the γ_1 coefficient on $(T_d * E_i)$ is negative in response to the question of whether the respondent supports one party rule. From the results in column (6) of Panel A, a one standard deviation increase in the treatment measure results in a 4 percentage point decrease in support for one-party rule. In Panel B, I also observe stronger support for democracy among the younger cohorts in more treated districts. Again, increasing the district-level treatment measure by one standard deviation results in an approximate 4 percentage point increase in support for democracy.

Table 2.7 reports results on perceptions of corruption, focusing on elected leaders, government officials and the police in Panels A, B and C respectively. Existing survey research on Tanzania indicates widespread perceptions of corruption in public officials (Chaligha et al, 2002). For example, in Panel C, 78 percent of respondents regarded corruption to be prevalent among the police service. For the preferred specification in column (6), the γ_1 coefficient is statistically significant and meaningful in magnitude. Increasing the district-level treatment measure from zero to 1 increased perception of corruption among MPs, government officials and the police by approximately 23 percent points, 30 percent points and 20 percent points respectively. The rejection of one-party rule, support for democracy today, and perception of widespread corruption among public officials is consistent with anecdotal evidence of disillusionment following the era of socialist planning, particularly following the interactions of peasant farmers with government bureaucrats and technical staff (von Freyhold, 1979; Fortmann, 1980).

Finally, in Table 2.8, I examine the outcome variable related to perceptions of ethnic versus national identity and trust of Tanzanians of other tribes. In Panel A, the outcome variable is a dummy which equals 1 if the respondent prefers national (i.e. Tanzanian iden-

TABLE 2.5: POLITICAL PARTICIPATION

	(1)	(2)	(3)	(4)	(5)	(6)
PANEL A						
Dep. Var: Dummy equal to one if respondent contacted local government official in the past year?						
Mean Dep. Var. = 0.279						
Young x Treatment	0.0954**	0.130***	0.154**	0.160**	0.159**	0.156**
	(0.0415)	(0.0448)	(0.0633)	(0.0645)	(0.0651)	(0.0633)
Observations	3,439	3,439	3,439	3,359	3,359	3,359
R-squared	0.114	0.116	0.117	0.117	0.117	0.117
PANEL B						
Dep. Var: Dummy equal to one if respondent contacted any government civil service official in the past year?						
Mean Dep. Var. = 0.109						
Young x Treatment	0.0669*	0.0985**	0.0949*	0.0914*	0.0988*	0.0887*
	(0.0397)	(0.0455)	(0.0497)	(0.0505)	(0.0502)	(0.0455)
Observations	3,434	3,434	3,434	3,354	3,354	3,354
R-squared	0.083	0.085	0.085	0.085	0.086	0.088
PANEL C						
Dep. Var: Dummy equal to one if respondent able to correctly name local MP?						
Mean Dep. Var. = 0.748						
Young x Treatment	0.219***	0.266***	0.220**	0.219**	0.234**	0.224**
	(0.0773)	(0.0794)	(0.0937)	(0.0979)	(0.101)	(0.0993)
Observations	1,864	1,864	1,864	1,838	1,838	1,838
R-squared	0.213	0.214	0.218	0.218	0.218	0.218
Geographic Controls	NO	YES	YES	YES	YES	YES
Demographic Controls	NO	NO	YES	YES	YES	YES
Agricultural Controls	NO	NO	NO	YES	YES	YES
Education Controls	NO	NO	NO	NO	YES	YES
Local Govt Controls	NO	NO	NO	NO	NO	YES

NOTES: Data is obtained from the Afrobarometer Tanzania Surveys: Rounds 3 and 4 data are used in Panels A and B; and Round 1 data for Panel C. Results are based on the regression model in equation (2.4). The coefficient on young x treatment (i.e. $T_d * E_i$) provides the difference-in-difference estimate comparing the mean outcome variable for older cohorts aged 48-78 years to a younger cohort aged 18-38 years. Geographic controls are for latitude, mean weighted slope and altitude. Demographic controls are for mean pre-treatment ethnolinguistic fragmentation, fraction of moslems and fraction of christians in the population based on the 1967 Tanzania Census. Agricultural controls are for per capita populations of cattle, sheep, and goats (Jensen and Mkama, 1968). Education control is for school enrollment rate in 1967 and local government controls are for per capita local government revenues in 1968-1970 (Jensen and Mkama, 1968). Robust standard errors are in parentheses and clustered at the 1978 district boundaries. *** p<0.01, ** p<0.05, * p<0.1.

TABLE 2.6: ONE PARTY RULE AND DEMOCRACY

	(1)	(2)	(3)	(4)	(5)	(6)
			PANEL A			
		Dep. Var: Dummy equal to one if respondent supports one party rule				
			Mean Dep. Var. = 0.429			
Young x Treatment	-0.0483	-0.115	-0.218*	-0.206	-0.222*	-0.209*
	(0.0968)	(0.111)	(0.122)	(0.126)	(0.126)	(0.125)
Observations	1,582	1,582	1,582	1,527	1,527	1,527
R-squared	0.175	0.177	0.180	0.176	0.179	0.180
			PANEL B			
		Dep. Var: Dummy equal to one if respondent supports democracy				
			Mean Dep. Var. = 0.644			
Young x Treatment	0.0108	0.0838	0.145	0.162	0.175	0.190*
	(0.0910)	(0.0916)	(0.104)	(0.106)	(0.106)	(0.107)
Observations	1,582	1,582	1,582	1,527	1,527	1,527
R-squared	0.219	0.222	0.224	0.224	0.226	0.227
Geographic Controls	NO	YES	YES	YES	YES	YES
Demographic Controls	NO	NO	YES	YES	YES	YES
Agricultural Controls	NO	NO	NO	YES	YES	YES
Education Controls	NO	NO	NO	NO	YES	YES
Local Govt Controls	NO	NO	NO	NO	NO	YES

NOTES: Data is obtained from Round 1 of the Tanzania Afrobarometer Survey. Results are based on the regression model in equation (2.4). The coefficient on young x treatment (i.e. $T_d * E_i$) provides the difference-in-difference estimate comparing the mean outcome variable for older cohorts aged 48-78 years to a younger cohort aged 18-38 years. Geographic controls are for latitude, mean weighted slope and altitude. Demographic controls are for mean pre-treatment ethnolinguistic fragmentation, fraction of moslems and fraction of christians in the population based on the 1967 Tanzania Census. Agricultural controls are for per capita populations of cattle, sheep, and goats (Jensen and Mkama, 1968). Education control is for school enrollment rate in 1967 and local government controls are for per capita local government revenues in 1968-1970 (Jensen and Mkama, 1968). Robust standard errors are in parentheses and clustered at the 1978 district boundaries. *** $p<0.01$, ** $p<0.05$, * $p<0.1$.

TABLE 2.7: PERCEPTIONS OF CORRUPTION

	(1)	(2)	(3)	(4)	(5)	(6)
PANEL A						
Dep. Var: Dummy equal to one if corruption is prevalent among elected leaders (MPs)?						
Mean Dep. Var. = 0.531						
Young x Treatment	0.165**	0.228**	0.212*	0.252**	0.227*	0.227*
	(0.0745)	(0.0932)	(0.117)	(0.120)	(0.134)	(0.135)
Observations	1,835	1,835	1,835	1,809	1,809	1,809
R-squared	0.174	0.176	0.177	0.171	0.171	0.171
PANEL B						
Dep. Var: Dummy equal to one if corruption is prevalent among government officials?						
Mean Dep. Var. = 0.597						
Young x Treatment	0.222**	0.248**	0.251*	0.304**	0.298**	0.300**
	(0.101)	(0.118)	(0.136)	(0.138)	(0.145)	(0.147)
Observations	1,835	1,835	1,835	1,809	1,809	1,809
R-squared	0.138	0.139	0.139	0.136	0.136	0.136
PANEL C						
Dep. Var: Dummy equal to one if corruption is prevalent among the police?						
Mean Dep. Var. = 0.784						
Young x Treatment	0.117*	0.172**	0.155	0.173*	0.189	0.196*
	(0.0617)	(0.0675)	(0.0978)	(0.103)	(0.119)	(0.116)
Observations	1,835	1,835	1,835	1,809	1,809	1,809
R-squared	0.108	0.112	0.113	0.110	0.110	0.110
Geographic Controls	NO	YES	YES	YES	YES	YES
Demographic Controls	NO	NO	YES	YES	YES	YES
Agricultural Controls	NO	NO	NO	YES	YES	YES
Education Controls	NO	NO	NO	NO	YES	YES
Local Govt Controls	NO	NO	NO	NO	NO	YES

NOTES: Data is obtained from the Afrobarometer Tanzania Surveys (Rounds 3 and 4). Results are based on the regression model in equation (2.4). The coefficient on young x treatment (i.e. $T_d * E_i$) provides the difference-in-difference estimate comparing the mean outcome variable for older cohorts aged 48-78 years to a younger cohort aged 18-38 years. Geographic controls are for latitude, mean weighted slope and altitude. Demographic controls are for mean pre-treatment ethnolinguistic fragmentation, fraction of moslems and fraction of christians in the population based on the 1967 Tanzania Census. Agricultural controls are for per capita populations of cattle, sheep, and goats (Jensen and Mkama, 1968). Education control is for school enrollment rate in 1967 and local government controls are for per capita local government revenues in 1968-1970 (Jensen and Mkama, 1968). Robust standard errors are in parentheses and clustered at the 1978 district boundaries. *** p<0.01, ** p<0.05, * p<0.1.

TABLE 2.8: PERCEPTIONS ON ETHNIC AND NATIONAL IDENTITY

	(1)	(2)	(3)	(4)	(5)	(6)
PANEL A						
Dep. Var: Dummy equal to one if respondent prefers national to ethnic identity						
Mean Dep. Var. = 0.753						
Young x Treatment	-0.0710	-0.121**	-0.167**	-0.170**	-0.173**	-0.163**
	(0.0568)	(0.0598)	(0.0700)	(0.0714)	(0.0719)	(0.0695)
Observations	3,396	3,396	3,396	3,316	3,316	3,316
R-squared	0.085	0.087	0.089	0.089	0.089	0.090
PANEL B						
Dep. Var: Dummy equal to one if respondent trusts Tanzanians of other tribes						
Mean Dep. Var. = 0.672						
Young x Treatment	0.0278	0.0614	0.0280	0.0229	0.0138	0.0211
	(0.0560)	(0.0603)	(0.0592)	(0.0592)	(0.0562)	(0.0550)
Observations	2,546	2,546	2,546	2,484	2,484	2,484
R-squared	0.102	0.106	0.107	0.107	0.108	0.108
Geographic Controls	NO	YES	YES	YES	YES	YES
Demographic Controls	NO	NO	YES	YES	YES	YES
Agricultural Controls	NO	NO	NO	YES	YES	YES
Education Controls	NO	NO	NO	NO	YES	YES
Local Govt Controls	NO	NO	NO	NO	NO	YES

NOTES: Data is obtained from the Tanzania Afrobarometer Survey. Panel A is based on results from Rounds 1, 3 and 4. Panel B is based on results from Rounds 1 and 3. Results are based on the regression model in equation (2.4). The coefficient on young x treatment (i.e. $T_d * E_i$) provides the difference-in-difference estimate comparing the mean outcome variable for older cohorts aged 48-78 years to a younger cohort aged 18-38 years. Geographic controls are for latitude, mean weighted slope and altitude. Demographic controls are for mean pre-treatment ethnolinguistic fragmentation, fraction of moslems and fraction of christians in the population based on the 1967 Tanzania Census. Agricultural controls are for per capita populations of cattle, sheep, and goats (Jensen and Mkama, 1968). Education control is for school enrollment rate in 1967 and local government controls are for per capita local government revenues in 1968-1970 (Jensen and Mkama, 1968). Robust standard errors are in parentheses and clustered at the 1978 district boundaries. *** p<0.01, ** p<0.05, * p<0.1.

tity) to ethnic identity. I find that the coefficient on the interaction $(T_d * E_i)$ is negative. Thus, the younger cohort in more villagized districts are more likely to state their identities in ethnic categories, rather than with the Tanzanian nation. Note that Tanzania is anecdotally viewed as an "ethnically" cohesive nation, and the mean for the national identity variable (approximately 0.75) is among the highest recorded in the Afrobarometer surveys for African nations. Yet, the results in Panel A appear initially surprising as the village formation process was also aimed at curbing ethnic loyalties and forging a greater sense of Tanzanian citizenship. Perhaps, in spite of the 1963 abolition of chieftaincies in Tanzania, ethnic loyalties remained in communities with large pre-Independence ethnic polities. If village formation was greater in such communities, then ethnic loyalties are likely to persist despite a history of high village formation in the 1970s.

2.6 Effects on Consumption, Community Participation and Public Goods Provision

2.6.1 OLS Results

In this section, I present OLS results on the impacts of the villagization program on current levels of consumption, community participation and public goods provision. My baseline estimating equation examines the relationship between the degree of villagization in an individual's district and the outcome variable, Y_{id}. This is presented in a regression framework as:

$$Y_{id} = \beta(T_d) + \delta_r + Z_d^{'}\phi + X_i^{'}\lambda + \epsilon_{id} \tag{2.5}$$

where Y_{id} refers to various outcome variables (log per capita consumption, participation in community activities or presence of school infrastructure), T_d refers to a historical measure of treatment intensity in district d, δ_r is a region fixed effect, and Z_d and X_i are respectively district and individual controls. In all specifications above, standard errors are clustered at the district level based on 1978 boundaries. When examining district-level outcomes, I run regressions similar to (2) above, but with district level outcomes. The coefficient of interest is β, which provides a point estimate of the relationship between villagization treatment and the outcome variable of interest. The individual level controls X_i' are for age, gender, and marital status; while the district controls are for pre-treatment characteristics such as geographic characteristics, demographic composition, agricultural output, economic activity, local government revenues, and health and education infrastructure. The OLS results based on equation (2.5) above are presented for household consumption, community participation and public goods provision in Tables 2.9 to 2.11.

The outcome variable in Table 2.9 is log of consumption per adult equivalent. I include region fixed effects in all columns, and thus examine whether variation in the treatment variable explains within-region differences in consumption. Column (1) presents the baseline specification, controlling only for geographic characteristics (namely latitude, slope, elevation and the mean and standard deviation of long-run precipitation) as well as the district-level of ethnic fragmentation. The results indicate that areas which experienced high levels of village formation are significantly worse off today. Increasing the district treatment intensity variable from zero to 1 results in approximately a 37.7 percent loss in per capita househod consumption today. I extend the analyses in columns (2) to (6) by introducing additional pre-1970 controls. In column (2), I control for district-level demographic characteristics by including the population density, fraction of the moslem population and the fraction of the Christian population. Agricultural controls are also introduced in column (3). The pre-1970 district agricultural characteristics provide information on the relative importance of agricultural employment across Tanzania, and also serve as proxies for pre-villagization district incomes. In column (3), I specifically include controls for the fraction of the population employed in agriculture, and the district per capita population of cattle, sheep, goats and donkeys. The point estimates increase in columns (2) and (3), and remains significant at the 5 percent level. To account for the availability of government services prior to 1970, I include controls for health and education infrastructure, and local government revenues in columns (4) and (5). Finally, in column (6), I also control for the relative support of the TANU party using district-level election outcomes from the 1970 Tanzanian presidential elections. The result in column (6) remains significant at the 10 percent level.

Overall, the adverse effects on household consumption reported in Table 2.9 is consistent with earlier and contemporary accounts which point to the welfare losses which occurred following village formation in Tanzania (Ergas, 1980; see Edwards, 2012 for recent discussion). The estimates of the consumption losses are also of comparable magnitude to estimates by Dippel (2011) on the welfare losses following the formation of Native American reservations in North America. Dippel estimates that forced integration of previously autonomous Indian bands into reservations lowered per capita incomes by about 30 percent today.

In Table 2.10, I examine outcomes on community participation. The results indicate that, conditional on region fixed effects, individuals born in more treated districts today report higher levels of attendance of local community meetings, and also are more likely to correctly name their village executive officer. These results remain robust to the inclusion of a variety of controls. In columns (5) and (6), I successively introduce pre-1970 controls for agricultural activity and for education enrollment. Thus, a 1 standard deviation increase in villagization treatment increases the probability of participating in village meetings today by about 6 percentage points, and also results in a 5 percentage point increase in ability to correctly name the local community leader.

Next, Table 2.11 examines the impact of treatment on provision of health and education infrastructure. Again, the treatment measure is associated with an increased probability of a community having a government primary school and a health facility. For example, a 1 standard deviation increase in treatment is associated with a 9 percentage point increase in the probability of having a school in the local community today. For both the availability of primary schools and local health facilities, the treatment effect remains significant across various specifications in columns (2) to (5) after including region fixed effects and various geographical and pre-1970 agricultural and economic controls.

TABLE 2.9: HOUSEHOLD PER CAPITA CONSUMPTION

	OLS (1)	OLS (2)	OLS (3)	OLS (4)	OLS (5)	OLS (6)
	Dep. Var: log per capita household consumption					
TreatmentIntensity	-0.377**	-0.434**	-0.414**	-0.386**	-0.354*	-0.305*
	(0.180)	(0.169)	(0.173)	(0.161)	(0.178)	(0.169)
Geographic controls	YES	YES	YES	YES	YES	YES
Demographic controls	NO	YES	YES	YES	YES	YES
Agricultural controls	NO	NO	YES	YES	YES	YES
Education and health controls	NO	NO	NO	YES	YES	YES
Local gov't revenues controls	NO	NO	NO	NO	YES	YES
TANU party (1970) controls	NO	NO	NO	NO	NO	YES
Region fixed effects	YES	YES	YES	YES	YES	YES
Observations	19,912	19,912	19,098	19,098	18,751	18,751
R-squared	0.165	0.184	0.195	0.197	0.201	0.202

*** $p<0.01$, ** $p<0.05$, * $p<0.1$. Data is obtained from the Tanzania Household Budget Survey (2000). All columns include region fixed effects, and control for ethnic fragmentation and geographic characteristics (latitude, slope, altitude, mean and standard deviation of long-term precipitation). Additional pre-treatment controls (pre-1970) are as follows: demographic controls (population density, fraction of moslem population and Christian population), agricultural controls (share of population employed in agriculture, per capita population of cattle, sheep, goats and donkeys), education controls (primary school enrollment rate), health controls (number of inhabitants per dispensary, number of hospital beds per capita), and local government revenues controls (per capita total local government revenue). TANU party (1970) controls are for 1970 electoral outcomes (fraction of population voting in favor of Nyerere/TANU and for voter turnout). Regression weighted using household sampling pweights provided in the Tanzania Household Budget Survey. Robust standard errors are in parentheses and clustered at the 1978 district boundaries.

TABLE 2.10: COMMUNITY PARTICIPATION

VARIABLES	OLS (1)	OLS (2)	OLS (3)	OLS (4)	OLS (5)	OLS (6)
Dep. var: Attendance of community meetings (mean depvar: 0.568)						
TreatmentIntensity	0.259***	0.178***	0.173***	0.221***	0.191***	0.171***
	(0.0499)	(0.0518)	(0.0542)	(0.0686)	(0.0555)	(0.0573)
Dep. var.: Ability to name village chief executive officer (mean = 0.328)						
TreatmentIntensity	0.236***	0.233***	0.234***	0.289***	0.289***	0.257***
	(0.0626)	(0.0803)	(0.0880)	(0.0743)	(0.0770)	(0.0806)
Geographic controls	YES	YES	YES	YES	YES	YES
Agricultural controls (1970)	NO	YES	YES	NO	YES	YES
Education controls (1970)	NO	NO	YES	NO	NO	YES
Region Fixed Effect	NO	NO	NO	YES	YES	YES
Observations	2,561	2,460	2,460	2,494	2,393	2,393

*** p<0.01, ** p<0.05, * p<0.1. Data is obtained from the community module of the Tanzania National Panel Survey (2008/9) conducted by NBS/World Bank. Region fixed effects are included in regressions for columns (4) to (6). All columns control for ethnic fragmentation and geographic characteristics (latitude, slope, altitude, mean and standard deviation of long-term precipitation). Agricultural controls are for: share of the district population employed in agriculture (from 1967 Tanzania National Census) and for per capita populations of cattle, sheep, goats and donkeys (from Jensen and Mkama, 1968). Other controls include pre-treatment primary education enrollment, share of population employed in manufacturing and local government revenues in 1967 (from Jensen and Mkama, 1968). Robust standard errors are in parentheses and clustered at the 1978 district boundaries.

TABLE 2.11: PUBLIC GOODS PROVISION

	(1) OLS	(2) OLS	(3) OLS	(4) OLS	(5) OLS
Dep. Var: Is there a government primary school within this community? (mean: .877)					
TreatmentIntensity	0.286***	0.380***	0.398***	0.450***	0.483***
	(0.0589)	(0.105)	(0.114)	(0.136)	(0.139)
Dep. Var: Is there a health facility in this community? (mean: 0.452)					
TreatmentIntensity	0.102	0.445***	0.367***	0.438***	0.379**
	(0.111)	(0.114)	(0.122)	(0.149)	(0.158)
Geographic controls	YES	YES	YES	YES	YES
Agricultural controls	NO	YES	YES	YES	YES
Employment, LGA revenues	NO	NO	YES	NO	YES
Region FE	NO	NO	NO	YES	YES

*** $p<0.01$, ** $p<0.05$, * $p<0.1$. Data is obtained from the community module of the Tanzania National Panel Survey (2008/9) conducted by NBS/World Bank. Region fixed effects are included in regressions for columns (4) and (5). All columns control for ethnic fragmentation and geographic characteristics (latitude, slope, altitude, mean and standard deviation of long-term precipitation). Agricultural controls are for: share of the district population employed in agriculture (from 1967 Tanzania National Census) and for per capita populations of cattle, sheep, goats and donkeys (from Jensen and Mkama, 1968). Other controls include pre-treatment primary education enrollment, share of population employed in manufacturing and local government revenues in 1967 (from Jensen and Mkama, 1968). Robust standard errors are in parentheses and clustered at the 1978 district boundaries.

2.6.2 Instrumental Variable Results

In this section, I consider the use of instrumental variables. The results in the previous section suggest that, conditional on various controls, there is a relationship between the treatment measure and current levels of household consumption, community participation and public goods provision. However, these results could still be biased if there are omitted variables correlated with the historical measure of villagization treatment. To address this problem, an instrument is needed which is correlated with historic treatment intensities, but uncorrelated with any other district or individual characteristics which may affect our outcome variables today. I consider an instrument motivated by two historical observations which influenced villagization: the incidence of prolonged droughts in the 1970s which coincided with the time period for implementation of villagization, and secondly, the variation in ethnolinguistic fragmentation across districts prior to village formation.

First, the incidence of droughts which occurred across parts of Tanzania between 1973-1975 is well documented in the historical literature (Nyerere, 1977). Scholars debated the importance of these droughts in explaining the government's economic policies and Tanzania's subsequent economic decline (Lofchie, 1978; Hyden, 1980). However, for the purposes of village formation, the droughts of the 1970s were important for two reasons. First, famine relief (such as imported maize) which was typically provided following such natural disasters would be selectively allocated to new village sites to encourage resettlement. Indeed, prior to the villagization campaigns, some initial resettlement was conducted following natural disasters, for example, when the Rufiji River flooded its banks in 1969 (Coulson, 1982, pp 247)[7]. Secondly, the droughts of the 1970s resulted in poor harvests which reduced the costs of relocation for many peasant farmers.

Goran Hyden (1980, pp 146-147), a noted Swedish scholar of this period, summarized the impact of the droughts on the villagization program:

> Villagization coincided with the drought that hit large parts of Tanzania in 1973/74. It is fair to say that from the viewpoint of the objectives of villagization, the drought was a blessing in disguise in that in many areas it facilitated

[7]Mamdani (1996), also documents a similar strategy adopted in neighbouring Mozambique in 1977-78 to villagize scattered rural households in Maputo and Gaza provinces following the flooding of the Limpopo and Incomati Rivers.

> the movement of people. There was little they left behind on the land and consequently it was easier for them to accept living in new villages. It is one of the reasons why some people did not object to moving but did so without government pressure...[W]eather continued to play havoc with agricultural production in some areas during 1974/75 and there was clearly a limit to what the peasants could produce before having settled in properly.

Below, I examine the variation in drought severity across Tanzania in the 1970s as an instrument for my IV/2SLS strategy. Using precipitation data covering 108 stations in Tanzania for the period 1960-2010 from the Tanzania Meteorological Agency, I examine the relationship between drought severity and the intensity of village formation in the 1970s. Maize served as the major food crop for most peasant communities (Herrick et al, 1968), and I focus on total precipitation recorded during the main planting season (*masika*) which approximately spans March to June each year. I compute a precipitation index for each district in two ways. First, I calculate a *ratio rainfall index* by dividing the total mean monthly precipitation during the *masika* seasons in 1973-1975 by the long-term average *masika* precipitation for the 50-year period, with this index censored at 1. A higher precipitation index means less severe drought. Second, I also examine a *standardized rainfall index* (Z-score) by subtracting the district long-term mean rainfall from the 1973-1975 mean, and dividing by the standard deviation, with this index censored at zero. The censoring of the drought measure is based on agricultural field experiments for maize which suggest minimal benefits from increasing rainfall above its long-run mean levels (see Hollinger and Chagnnon, 1993; and Dell, 2011). Figure 2.3 displays the variation in drought severity across Tanzania, based on the ratio rainfall index for the period 1973-75.

Besides the droughts of the 1970s, ethnic and linguistic diversity hindered communication and cooperation, so that village formation was likely to be more successful in less ethnically fragmented districts (McHenry, 1979, pp 175-6)[8]. I calculate a measure of eth-

[8]von Freyhold (1979, pp 142) also provides an example of such an encounter between two neighboring ethnic groups in Ilala district attempting to move into a common village during the ujamaa period prior to the national villagization program: "According to the Tibili villagers it was the Tibili site that was intended to be the site of the Ujamaa village all along. According to the Chanika villagers it was the Chanika site, and it was only when they went to cut poles to start house building that they realized that the village was to be at Tibili, for the government lorry dropped the poles at the Tibili site. When the villagers complained, there was an argument during which the divisional executive officer announced that Tibili was the approved site and that those who were not interested in building at Tibili could stay out."

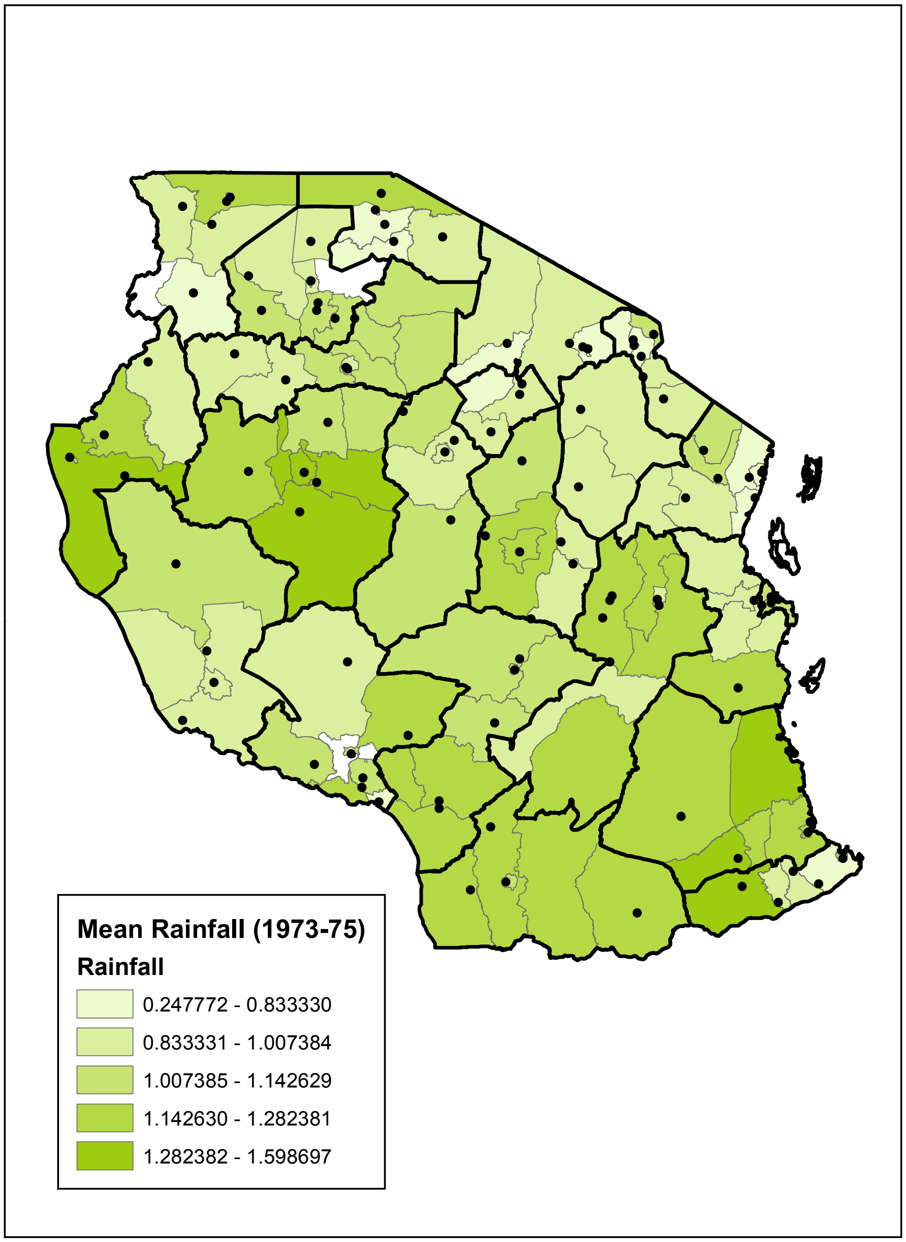

Figure 2.3: **Map of Drought Severity Index for 1973-75**. The mean rainfall for 1973-75 *masika* seasons is divided by the corresponding long-run average. Region boundaries are darkened. Black dots point to location of rainfall stations.

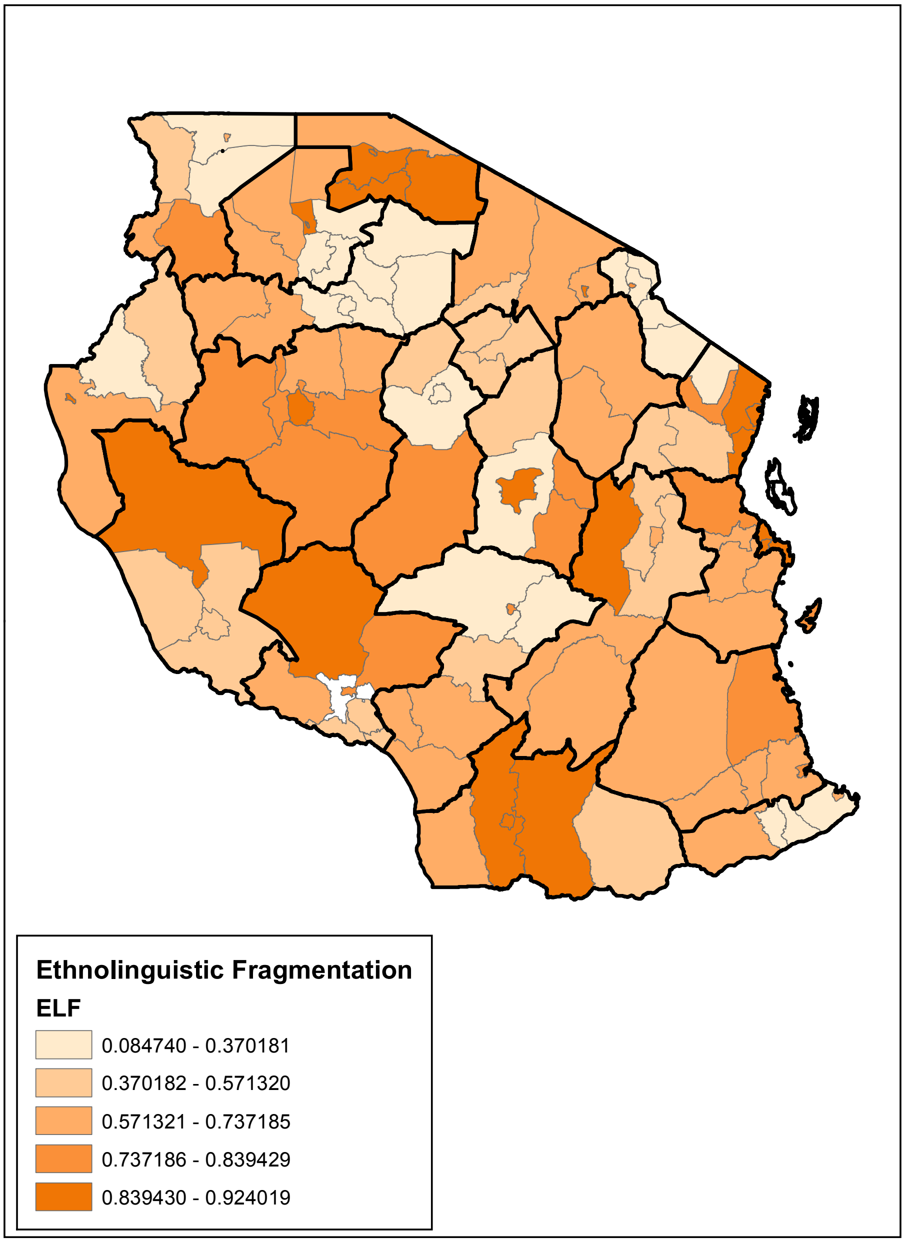

Figure 2.4: **Map of Ethnolinguistic fragmentation,** ELF_d (Based on 1957 Tanzania National Census). The district-level measure of ethnolinguistic fractionalization is computed as $ELF_d = (1 - \sum e_{id}^2)$, where e_i is the fraction of the population in district, d, belonging to ethnic group, i. Region boundaries are darkened.

nolinguistic fragmentation, ELF_d, using the 1957 Tanzania National Census (prior to the Arusha Declaration), which is the most recent national census with data on ethnic populations at the district level. The district-level measure of ethnolinguistic fractionalization is computed using the usual formulation in the literature, $ELF_d = (1 - \sum e_{id}^2)$, where e_i is the fraction of the population in district, d, belonging to ethnic group, i (Mauro, 1995; Posner, 2004).[9][10] Figure 2.4 displays the variation in ELF_d across Tanzanian districts in 1957. The historic levels of ELF_d is potentially correlated with the error term in equation (2.5), and thus does not satisfy the exclusion restrictions. I therefore consider the interaction of the drought and ethnolinguistic fragmentation as an instrument, while controlling for the direct effect of ELF.

Specifically, I examine two possible first stage regressions. First, using the drought measure as an instrument, and next, using the interaction of the drought and ELF while controlling for the level of ELF. The first-stage regressions are stated below:

$$T_d = \gamma_{0a} + \gamma_{1a} Rainfall_d + X_d^{'} \lambda_a + \epsilon_d \tag{2.6}$$

$$T_d = \gamma_{0b} + \gamma_{1b} Rainfall_d * ELF_d + \gamma_2 ELF_d + X_d^{'} \lambda_b + \phi_d \tag{2.7}$$

where (T_d) is the treatment measure, X_d is a vector of controls including geographic characteristics and district pre-treatment characteristics. The geographic controls include controls for latitude, slope, altitude, and the mean and standard deviation of precipitation over the period 1960-2010. We expect the coefficients $\gamma_{1a} < 0$ and $\gamma_{1b} < 0$ since both higher rainfall and higher ELF result in less village formation.

Table 12 reports results of the first-stage relationship. Columns (1) and (2) present results with the ratio of mean precipitation to the long run mean, while columns (3) and (4) utilize the standardized precipitation index (Z-score). Columns (1) and (3) control only for geographical characteristics (namely: latitude, slope, altitude and the mean and standard deviation of long-run precipitation), while in columns (2) and (4), I include additional

[9]This measure represents the probability that any two randomly chosen individuals in district, d, belong to different ethnic groups.

[10]The district-level ethnic composition appears relatively stable in the early independence period: for example, the correlation between $ELF_{d,1957}$ and $ELF_{d,1948}$ is 0.978.

controls for pre-1970 agricultural characteristics and education enrollment. In Panel A, the instrument used is only the rainfall measure, while in Panel B, I use the interaction of the rainfall measure with ELF. In all specifications in both Panels A and B, there is a significant and negative relationship between the instrument and the intensity of villagization. The instruments are however weak (F-statistics ranging between 6.00 and 9.4), and thus suggesting that the 2SLS estimates are likely to be biased towards my OLS estimates (Bound, Jaeger, and Baker 1995; Staiger and Stock 1997).

The validity of the instrument rests on two conditions (Angrist and Pischke, 2008). First, the existence of a first stage relationship showing the correlation of the instrument with the endogenous regressor (village treatment intensity). This is shown in Table 2.12 above. The second condition for the IV strategy is the exclusion restriction, which requires that the instrument be uncorrelated with other determinants of the outcome measure of interest. The exclusion restriction is comprised of two parts: first, that the instrument be as good as randomly assigned conditional on covariates, and second, that the instrument affects outcome variables today only via the village formation channel.

The exclusion restrictions cannot be directly tested, but I conduct two further falsification exercises as further checks of the validity of the instrument. In particular, it is necessary to check that village formation is not simply occurring in drought-prone areas, but rather occurs specifically in areas which experienced droughts during the stipulated time period for villagization of 1973-75. In the falsification exercises, I examine placebo droughts which may have occurred in the three-year period just prior to (1970-1972), and just following (1976-1978), the time window of the implementation of the villagization policy. The results of the placebo checks are reported in Table 2.13. Although the point estimates are still negative, I do not find a significant relationship between drought severity and the intensity of villagization for the specifications in Panels A and B of Table 2.13.

Tables 2.14-2.16 report results for the 2SLS regressions for log equivalent consumption, community participation and public goods provision outcomes using the rainfall instrument only. The results for consumption are presented in Table 2.14. Columns (1)-(3) repeat OLS results while columns (4)-(6) present the 2SLS results. I report LIML results in columns (7)-(9) as it may provide better finite-sample properties in the case of weak instruments. I also report Anderson-Rubin confidence intervals, which are robust to weak instruments, in

TABLE 2.12: FIRST STAGE REGRESSIONS

	Percent ofmean **(1973-75)** (1)	Percent of mean **(1973-75)** (2)	Z-score **(1973-75)** (3)	Z-score **(1973-75)** (4)
PANEL A				
Rainfall	-0.528**	-0.709**	-0.191**	-0.236**
	(0.253)	(0.281)	(0.0818)	(0.0918)
	[0.261]	[0.258]	[0.085]	[0.086]
Geographic controls	YES	YES	YES	YES
Agric. and education controls (1970)	NO	YES	NO	YES
F-statistic:	4.37	6.37	5.43	6.63
Observations	82	79	82	79
R-squared	0.319	0.398	0.325	0.401
PANEL B				
Rainfall * Ethnolinguistic Fragmentation	-0.993**	-1.257***	-0.353***	-0.415***
	(0.403)	(0.413)	(0.129)	(0.135)
	[0.410]	[0.352]	[0.134]	[0.121]
Geographic controls	YES	YES	YES	YES
Agric. and education controls (1970)	NO	YES	NO	YES
F-statistic:	6.08	9.28	7.43	9.42
Observations	82	79	82	79
R-squared	0.334	0.416	0.341	0.418

*** p<0.01, ** p<0.05, * p<0.1. Table based on monthly precipitation data from the Tanzania Meteorological Agency for 108 rainfall stations over the period 1960-2010. I calculate the rainfall index over the period 1973-75 in 2 ways: first, using mean total precipitation during the main (*masika*) rainy season divided by the corresponding long-term average, with the index censored at 1. Or secondly, using a standardized rainfall index (Z-score) with the index censored at 0. A higher rainfall index means less severe drought. In Panel A the insturment is only drought severity. In Panel B, the instrument refers to the interaction of the rainfall index and ethnolinguistic fragmentation (ELF). All columns control for the direct effect of ELF. Geographic controls are for latitude, slope, altitude, mean and standard deviation of long-run precipitation. Robust standard errors are in parentheses. Conley errors are in square brackets.

TABLE 2.13: PLACEBO FIRST STAGE REGRESSIONS

	Percent of mean **(1970-72)** (1)	Percent of mean **(1970-72)** (2)	Percent of mean **(1976-78)** (3)	Percent of mean **(1976-78)** (4)
PANEL A				
Rainfall	-0.162	-0.119	-0.443	-0.464
	(0.198)	(0.233)	(0.404)	(0.461)
	[0.185]	[0.208]	[0.430]	[0.492]
Geographic controls	YES	YES	YES	YES
Agric. and education controls (1970)	NO	YES	NO	YES
Observations	78	75	77	74
R-squared	0.232	0.304	0.266	0.359
PANEL B				
Rainfall * Ethnolinguistic Fragmentation	-0.542	-0.504	-0.483	-0.379
	(0.328)	(0.436)	(0.621)	(0.657)
	[0.307]	[0.408]	[0.645]	[0.694]
Geographic controls	YES	YES	YES	YES
Agric. and education controls (1970)	NO	YES	NO	YES
Observations	78	75	77	74
R-squared	0.252	0.319	0.262	0.352

*** $p<0.01$, ** $p<0.05$, * $p<0.1$. Table based on monthly precipitation data from the Tanzania Meteorological Agency for 108 rainfall stations covering the period 1960-2010. Rainfall index calculated for the three-year period prior to villagization (1970-72) and following villagization (1976-78). Rainfall index is calculated as the mean total precipitation during the main (*masika*) rainy season divided by the corresponding long-term average, with the index censored at 1. In Panel A, the instrument is only the Rainfall index. In Panel B, the instrument refers to the interaction of the rainfall index and ethnolinguistic fragmentation (ELF). All columns control for the direct effect of ELF. Geographic controls are for latitude, slope, altitude, mean and standard deviation of long-run precipitation. Robust standard errors are in parentheses. Conley errors are in square brackets.

square brackets in Table 2.14. The 2SLS and LIML results indicate even larger magnitudes for consumption losses compared to the previous OLS results. The increase in magnitude of the point estimates in the IV regressions may be because the instrument, in this instance, provides a local average treatment effect (LATE) of the impact of villagization treatment on consumption for those districts (compliers) in which village formation was particularly induced by the drought.

For the results on attendance of community meetings in Table 2.15, the 2SLS estimates yields a treatment effect with a point estimate of 0.66 (s.e.=0.25) in the preferred regression in column (6). This result includes controls for geographic characteristics and pre-treatment agricultural, health and education characteristics. Thus a 1 standard deviation increase in treatment intensity measure resulted in about a 12 percentage point increase in participation rates in community meetings. Similarly, the size of the point estimates for the treatment effect for the ability to name the local village official similarly remains positive, although the IV estimates are now not precisely estimated. In Table 2.16, the size of the estimated 2SLS treatment effect for the primary school outcome variable is positive and significant in all specifications, and larger in magnitude than the corresponding OLS estimates. For example, for the primary school outcomes in Table 2.16, the probability of having a government primary school increases from about 0.398 (s.e.=0.114) in the OLS regression in column (3), to 0.795 (s.e.=0.401) in the IV regression in column (6). The 2SLS results for the health facility outcome variable are however not precisely estimated.

TABLE 2.14: HOUSEHOLD PER CAPITA CONSUMPTION

VARIABLES	OLS			2SLS			LIML		
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
				Dep. Var: log per capita household consumption					
TreatmentIntensity	-0.461***	-0.386***	-0.468***	-0.778**	-0.842**	-0.826*	-0.778**	-0.844**	-0.836*
	(0.0936)	(0.112)	(0.100)	(0.314)	(0.352)	(0.443)	(0.314)	(0.354)	(0.458)
Anderson-Rubin				[-.951, -.606]	[-1.089, -.583]	[-.859, -.398]	[-.951, -.606]	[-1.089, -.583]	[-.859, -.398]
Geographic controls	YES	YES	YES	YES	YES	YES	YES	YES	YES
Agricultural controls (1970)	NO	YES	YES	NO	YES	YES	NO	YES	YES
Education controls (1970)	NO	NO	YES	NO	NO	YES	NO	NO	YES
Observations	19,871	19,058	19,058	19,266	18,453	18,453	19,266	18,453	18,453
R-squared	0.104	0.111	0.123	0.096	0.098	0.118	0.096	0.098	0.117

*** $p<0.01$, ** $p<0.05$, * $p<0.1$. Data is obtained from the 2000 Tanzania Household Budget Survey. Columns (1) to (3) repeat OLS regressions, and Columns (4) to (9) provide instrumental variable estimates using the ratio rainfall index (ratio of 1974-1976 mean precipation to long-term average, with index censored at 1) as instrument. All columns control for ethnic fragmentation and geographic characteristics (latitude, slope, altitude, mean and standard deviation of long-term precipitation). Agricultural controls are for: share of the district population employed in agriculture (1967 Census) and for per capita populations of cattle, sheep, goats and donkeys (Jensen and Mkama, 1968). Education control is for school enrollment rate in 1967. Robust standard errors are in parentheses and clustered at the 1978 district boundaries. Anderson-Rubin confidence intervals are provided in square brackets.

TABLE 2.15: COMMUNITY PARTICIPATION

VARIABLES	OLS			2SLS			LIML		
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
	Dep. var: Attendance of community meetings (mean depvar: 0.568)								
TreatmentIntensity	0.259***	0.178***	0.173***	0.861**	0.543**	0.660***	0.861**	0.553**	0.675***
	(0.0499)	(0.0518)	(0.0542)	(0.402)	(0.239)	(0.251)	(0.402)	(0.246)	(0.259)
Observations	2,561	2,460	2,460	2,494	2,393	2,393	2,494	2,393	2,393
R-squared	0.092	0.105	0.107	0.031	0.090	0.084	0.031	0.089	0.083
	Dep. var.: Ability to name village chief executive officer (mean = 0.328)								
TreatmentIntensity	0.236***	0.233***	0.234***	0.654	0.464	0.515	0.654	0.479	0.534
	(0.0626)	(0.0803)	(0.0880)	(0.474)	(0.338)	(0.352)	(0.474)	(0.358)	(0.373)
Observations	2,564	2,463	2,463	2,497	2,396	2,396	2,497	2,396	2,396
R-squared	0.054	0.074	0.075	0.026	0.059	0.058	0.026	0.058	0.057
Geographic controls	YES	YES	YES	YES	YES	YES	YES	YES	YES
Agricultural controls (1970)	NO	YES	YES	NO	YES	YES	NO	YES	YES
Education controls (1970)	NO	NO	YES	NO	NO	YES	NO	NO	YES

*** $p<0.01$, ** $p<0.05$, * $p<0.1$. Data is obtained from the Tanzania National Panel Survey (2008/9) conducted by NBS/World Bank. Columns (1) to (3) repeat OLS regressions, and Columns (4) to (9) provide instrumental variable estimates using the ratio of 1974-1976 mean precipation to long-term average as instrument. All columns control for ethnic fragmentation and geographic characteristics (latitude, slope, altitude, mean and standard deviation of long-term precipitation). Agricultural controls are for: share of the district population employed in agriculture (1967 Census) and for per capita populations of cattle, sheep, goats and donkeys (Jensen and Mkama, 1968). Education control is for school enrollment rate in 1967. Robust standard errors are in parentheses and clustered at the 1978 district boundaries.

TABLE 2.16: PUBLIC GOODS PROVISION

VARIABLES	OLS (1)	OLS (2)	OLS (3)	2SLS (4)	2SLS (5)	2SLS (6)
Dep. Var: Is there a government primary school within this community? (mean: .877)						
TreatmentIntensity	0.286***	0.380***	0.398***	0.441**	0.797*	0.795**
	(0.0589)	(0.105)	(0.114)	(0.186)	(0.423)	(0.401)
Dep. Var: Is there a health facility in this community? (mean depvar: 0.452)						
TreatmentIntensity	0.102	0.445***	0.367***	-0.305	-0.0725	-0.0975
	(0.111)	(0.114)	(0.122)	(0.214)	(0.429)	(0.387)
Geographic controls	YES	YES	YES	YES	YES	YES
Agricultural controls (1970)	NO	YES	YES	NO	YES	YES
Education controls (1970)	NO	NO	YES	NO	NO	YES

*** p<0.01, ** p<0.05, * p<0.1. Data is obtained from the community module of the Tanzania National Panel Survey (2008/9) conducted by NBS/World Bank. Columns (1) to (3) repeat the OLS regressions in Table 6 for comparison. Columns (4) to (6) provide instrumental variable estimates using the ratio of 1974-1976 mean precipation to long-term average as instrument. All columns control for ethnic fragmentation and geographic characteristics (latitude, slope, altitude, mean and standard deviation of long-term precipitation). Agricultural controls are for: share of the district population employed in agriculture (1967 Census) and for per capita populations of cattle, sheep, goats and donkeys (Jensen and Mkama, 1968). Education control is for school enrollment rate in 1967. Robust standard errors are in parentheses and clustered at the 1978 district boundaries.

2.7 Channels of Persistence

The discussion thus far presents villagization as a package of institutional reforms in the Tanzanaian countryside which has resulted in long-term effects on education, household consumption and local community participation. In this section, I attempt to unpack the specific components of the reform program which may be important. I provide some preliminary evidence to explain the decline in consumption and the increased availability of local public goods.

2.7.1 Consumption and Economic Activities

The OLS results presented earlier indicate an approximate 38 percent decline in consumption today in moving the district treatment measure from zero to 1. The IV results indicate even more severe declines in consumption. A possible explanation for the persistence of the consumption losses is the decline in agriculture and the relative lack of economic diversification in areas which experienced large village formation in the 1970s. Putterman (1995a) observes that although the villagization program increased access of villages to government administration and local public goods, it increased distances to farmlands. The withdrawal of government transportation and marketing sevices in the 1980s increased transportation costs for peasant farmers in less accessible areas, resulting in a decline in cash crop farming.[11]

There has been a notable lack of economic diversification away from peasant agriculture in many rural districts in Tanzania following the villagization campaigns of the 1970s (see Collier et al, 1986, Chapter 6). I am able to shed some light on the lack of economic diversification by examining a differences-in-differences regression across various year-of-birth cohorts using microdata from the 2002 Tanzania National Census. I focus here on the respondent's employment activity, and examine the fraction of the population which transitions away from agriculture into manufacturing, construction services, trading activity

[11] Future work would possibly examine more closely the impact of villagization on access to rural transportation infrastructure. There are detailed cartographic records prior to, and following, villagization which should permit a detailed exercise of settlement patterns for some regions in Tanzania.

or public administration. As in section 2.5, I examine a regression of the form:

$$Y_{idcs} = \beta(Young_c \times TreatmentIntensity_d) + \delta_d + \delta_c + ((Young_c \times X_d^{'})\lambda + \epsilon_{idc} \quad (2.8)$$

where Y_{idc} is a dummy for whether individual i, in district d, and year-of-birth cohort, c, is employed in sector, s; the dummy $Young_c = 1$ if year of birth occured after 1968 (and thus likely to have been exposed to villagization), and δ_d and δ_c are respectively district and birth cohort fixed effects. X_d is a vector containing various district-level controls such as time-invariant geographic controls. In robustness checks, I control for other pre-treatment characteristics such as pre-1970 primary school enrollment and health infrastructure, which enter into the equation above interacted with the dummy for $Young_{dc}$.

The results from equation (2.8) above are presented in Table 2.17. I report results on the coefficient for *Young* $\times$ *TreatmentIntensity* in Panel A, and for male and female subsamples in Panels B and C respectively. The results in column (1) of Panel A, indicate an increase in government public employment, suggesting that the central and local government bureaucracy remain sources of employment in previously high treatment areas. Employment in trading activities (retail and wholesale trade) decline, and this effect is significant in the results reported in Panel A. The decline in trading activity is driven by the effect for the male subsample (see Panel B). The effects for transitioning to employment in manufacturing and construction services are not precisely estimated, but decline (with significant point estimates) in the female subsample (see Panel C).

2.7.2 Public Goods Provision

The OLS results in Section 2.6.1 also point to an increased availability of local public goods in villages which experienced a high level of treatment villages. The creation of new villages in the 1970s in Tanzania in many cases involved the integration of diverse ethnic and religious groups into new communities. Evidence from the recent literature suggests that ethnic heterogeneity may inhibit collective action and thus lower economic outcomes and the provision of local public goods (Alesina et al. 1999; Alesina and La-Ferrara 2000; Miguel and Gugerty 2005). However, if village formation was higher in more ethnicially homegeneous districts, then we can expect high public good provision in high treatment areas today. The results reported in Section 2.6.1 already include controls for pre-villagization ethnolinguistic fragmentation, suggesting that additional mechanisms may be at work.

The historical account suggests two channels which may be important in explaining public good provision today, namely: electoral participation (Cliffe, 1967; Barker and Saul, 1974) and the development of local government fiscal capacities (Fortmann, 1980; Ngware and Haule, 1992). First, the widespread mobilization in developmental villages in the late 1970s raised the political consciousness of many Tanzanian citizens, and encouraged high levels of participation in national elections. In particular, areas which historically had a high fraction of the population in development villages tend to support the ruling TANU/CCM party which has dominated national and local politics in Tanzania. Support for the ruling party, may therefore be one channel which explains the relatively higher availability of public goods in high treatment districts today.

Second, the developmental villages introduced some form of (informal) taxation among village residents for most of the 1970s. Such taxation occurred either via working on village farms, participating in self-help communal activities to maintain village infrastructure such as granaries, or selling produce to village cooperatives (Fortmann, 1980; Collier, Radwan and Wangwe, 1986). The poll tax introduced by British colonial officials proved highly unpopular and had been abolished by Nyerere's TANU government in the years following Independence. Thus besides the implicit taxation of agricultural produce (by state marketing co-operatives), the informal taxation imposed by village councils constituted the next major form of taxation imposed on peasants in the countryside. Households and officials in highly villagized districts therefore historically obtained greater experience in taxation, and it is plausible that these effects may persist today.

In Table 2.18, I present some initial results exploring the electoral participation and local government tax channel. Tanzania remained a one-party state from 1960 to 1992, and the nationalist TANU/CCM party dominates both parliamentary and presidential elections in Tanzania. I therefore use the combined share of votes received by opposition parties as a measure of electoral competition. In Panel A of Table 2.18, I report results using data from the 2010 presidential elections. Conditional on region fixed effects, I find that treated districts are associated today with a reduced vote share for opposition parties. The preferrred specification is provided in column (6) in which I include controls for the pre-1970 vote share of the TANU party, as well as time-invariant geographic controls and various pre-1970 district characteristics. The magnitude of the point estimate indicates that increasing district-level treatment variable from zero to 1 results in a 13 percentage point decrease in the opposition vote share. The results in Panel A suggest that treated districts which experienced high levels of government-planned villagization in the late 1970s remain loyal supporters of the governing TANU/CCM party. In a model of distributive politics in which elected officials reward their loyal bases (Cox and McCubbins, 1986), then the continued allegiance of treated districts to the TANU/CCM may partly explain the higher provision of public goods available in these districts. In Panel B, I provide some preliminary evidence on current local government revenues derived from taxing agricultural produce (i.e. the produce cess). The outcome variable is the per capita agricultural taxes collected in units of Tanzanian shillings, and normalized by the fraction of the population employed in agriculture. In the specification in column (6), I find that a one standard deviation increase in the treatment intensity measure increases per capita produce cess collected by about 78 Tanzanian Shillings today.

TABLE 2.17: CHANGES IN OCCUPATIONS

	(1)	(2)	(3)	(4)
PANEL A: FULL SAMPLE				
Employed in manufacturing	-0.000704	-0.000656	-0.00279	0.00131
	(0.00463)	(0.00464)	(0.00608)	(0.00264)
Employed in construction industry	0.000463	0.000514	0.000660	0.000981
	(0.000999)	(0.00100)	(0.00107)	(0.00112)
Employed in trading	-0.0150**	-0.0146**	-0.0137*	-0.0154*
	(0.00669)	(0.00642)	(0.00719)	(0.00781)
Employed in public administration and other services	0.0430***	0.0430***	0.0412***	0.0420***
	(0.00467)	(0.00469)	(0.00518)	(0.00561)
PANEL B: MALE SUBSAMPLE				
Employed in manufacturing	-0.00108	-0.00105	-0.00588	0.00521
	(0.00965)	(0.00969)	(0.0134)	(0.00456)
Employed in construction industry	-0.000879	-0.000809	-0.000199	0.00129
	(0.00217)	(0.00218)	(0.00242)	(0.00244)
Employed in trading	-0.0356***	-0.0355***	-0.0335***	-0.0348***
	(0.00741)	(0.00736)	(0.00829)	(0.00973)
Employed in public administration and other services	0.0551***	0.0551***	0.0540***	0.0585***
	(0.00698)	(0.00701)	(0.00833)	(0.00866)
PANEL C: FEMALE SUBSAMPLE				
Employed in manufacturing	-0.00551***	-0.00548***	-0.00557***	-0.00590***
	(0.00171)	(0.00169)	(0.00184)	(0.00197)
Employed in construction industry	-0.00110***	-0.00111***	-0.00118***	-0.00128***
	(0.000382)	(0.000377)	(0.000379)	(0.000391)
Employed in trading	0.00170	0.00265	0.00286	0.000246
	(0.00782)	(0.00734)	(0.00817)	(0.00834)
Employed in public administration and other services	0.0199***	0.0199***	0.0177***	0.0181***
	(0.00501)	(0.00494)	(0.00498)	(0.00584)
Young dummy * primary enrollment rate in 1970	NO	YES	YES	YES
Young dummy * health infrastructure in 1970	NO	NO	YES	YES
Young dummy * other pre-treatment district characterstics	NO	NO	NO	YES

*** p<0.01, ** p<0.05, * p<0.1. Sectoral classifications are as follows: agriculture (farming, fishing and forestry), manufacturing (mining, manufacturing, electricity, gas and water), construction industry (construction services), trading (wholesale and retail trade) and public administration and other services (public administration (80 percent) and defense, transportation and communications, financial services and insurance). Results are presented for the coefficient on *Young * TreatmentIntensity*. The difference-in-difference model compares outcomes for a younger cohort born after 1968 (and thus exposed to villagization) versus an older cohort born prior to 1960. Data is obtained from IPUMS microsample of 2002 Tanzania Census. All columns include fixed effects for year of birth cohorts and districts. Controls enter the regression interacted with the young dummy. Column 2 controls for pre-treatment primary enrollment in 1970. Column 3 includes controls for health infrastructure in 1970. Column 4 includes various geographic and demographic controls (for latitude, slope, altitude, ethnic fragmentation, share of christians, livestock populations, and local government revenues — see Jensen and Mkama, 1968). Robust standard errors are in parentheses and clustered at the 1978 district boundaries.

TABLE 2.18: CHANNELS OF PERSISTENCE

	OLS (1)	OLS (2)	OLS (3)	OLS (4)	OLS (5)	OLS (6)
PANEL A						
Dep. Var: Opposition Vote Share (mean = 0.35)						
TreatmentIntensity	-0.0251	-0.0457	-0.102*	-0.118**	-0.131**	-0.131**
	(0.0665)	(0.0668)	(0.0534)	(0.0529)	(0.0524)	(0.0545)
Observations	110	110	106	106	104	104
R-squared	0.576	0.621	0.799	0.799	0.821	0.824
PANEL B						
Dep. Var: Produce cess per capita (mean = 454.7 TSh)						
TreatmentIntensity	421.8	554.1**	636.5**	384.4	490.5*	393.8*
	(270.6)	(267.7)	(289.7)	(259.6)	(283.8)	(225.7)
Observations	105	105	102	102	100	100
R-squared	0.505	0.559	0.663	0.673	0.641	0.651
Geographic controls	YES	YES	YES	YES	YES	YES
Demographic controls	NO	YES	YES	YES	YES	YES
Agricultural controls	NO	NO	YES	YES	YES	YES
Education and health controls	NO	NO	NO	YES	YES	YES
Local gov't revenues controls	NO	NO	NO	NO	YES	YES
TANU party (1970) controls	NO	NO	NO	NO	NO	YES
Region fixed effects	YES	YES	YES	YES	YES	YES

*** $p<0.01$, ** $p<0.05$, * $p<0.1$. In Panel A, dependent variable is the share of the vote in a given district received by opposition parties in the 2005 Tanzanian Presidential Elections. Electoral results are obtained from the National Electoral Commission of Tanzania. In Panel B, the dependent variable is the per capita agricultural taxes collected (produce cess) in units of Tanzanian Shillings, and normalized by the share of district population employed in agriculture in the 2002 Tanzanian Agricultural Census. Data on local government tax revenues is obtained from the Tanzanian Prime Ministers Office Regional Administration and Local Government (PMO-RALG, TAMISEMI). All columns include region fixed effects, and control for ethnic fragmentation and geographic characteristics (latitude, slope, altitude, mean and standard deviation of long-term precipitation). Additional pre-treatment controls (pre-1970) are as follows: demographic controls (population density, fraction of moslem population and Christian population), agricultural controls (share of population employed in agriculture, per capita population of cattle, sheep, goats and donkeys), education controls (primary school enrollment rate), health controls (number of inhabitants per dispensary, number of hospital beds per capita), and local government revenues controls (per capita total local government revenue). TANU party (1970) controls are for 1970 electoral outcomes (fraction of population voting in favor of Nyerere/TANU and for voter turnout). Robust standard errors are in parentheses and clustered at the 1978 district boundaries.

2.8 Conclusions

Combining historic data from various years in the 1960s and 1970s with contemporary datasets, this Chapter has provided evidence of the impact of Tanzania's villagization program on contemporary development outcomes. Using IPUMS microdata from the 1988 Tanzania national census, I first documented that districts with a high fraction of the population living in development villages increased educational outcomes (primary school completion rates, literacy rates and total years of schooling) for both males and females. In addition, using results from the Afrobarometer survey, I provide some evidence that a higher level of the treatment variable is associated today with greater political participation and greater support for democracy, but a rejection of one-party rule and worse perception of corruption among government officials. Next, using various pre-1970 district controls and recent household surveys, I report OLS results which confirm the decline in household consumption in areas which experienced a high intensity of the villagization treatment measure. I also present results on community participation and the availability of some local public goods. To address the endogeneity in village formation, I exploit the incidence of droughts in the 1970s which facilitated the re-settlement of peasants to provide instrumental variable estimates.

Finally, I provide some preliminary evidence on possible channels of persistence. To account for the severe consumption losses reported, I document that economic diversification (i.e. a transition away from peasant agriculture) has been limited in areas which historically experienced high levels of villagization. I present additional evidence that, conditional on various controls, the treatment measure is correlated with lower oppostion vote shares today. This suggests that districts which experienced high levels of villagization in the 1970s remain loyal supporters of the ruling TANU/CCM party, partly explaining the higher availability of public goods in these districts today. The experience of late 1970s villagization transformed the Tanzanian countryside, and the results in this preliminary project suggest that the institutional developments during this period have persistent effects across Tanzanian districts today.

Chapter 3

Penicillin and Yaws Control: Some Evidence from Ghana

3.1 Introduction

The recent focus of donor agencies and the public health community on tackling three major diseases — tuberculosis, HIV/AIDS and malaria — has resulted in a reduced priority being accorded to other widespread tropical diseases. These illnesses — commonly referred to as "neglected tropical diseases" — receive only about 0.6 percent of global health development assistance but afflict nearly 96 percent of the "bottom billion" of the world's population who may not have HIV (Liese and Schubert, 2009). The list of these diseases varies, but often includes: blinding trachoma, Buruli ulcer, Chagas disease, dengue, dracunculiasis, human African trypanosomiasis, Japanese encephalities, leishmaniases, leprosy, lymphatic filiarisis, onchocerciasis, schistosomiases, soil-transmitted helminthiasis and yaws. The neglect partly arises because existing methods of prioritizing disease severity rely on the calculation of DALYs (disability adjusted life years) which are heavily weighted by mortality rates. Neglected tropical diseases, on the other hand, tend to be rarely fatal and consequently are often not highly prioritized. The treatment of these diseases, however, is often highly cost-effective and successful control could also provide other economic benefits such as higher education attainment and increased worker productivity (Canning, 2006). Public health experts argue that neglected tropical diseases constitute the “low hanging fruit” diseases, and could also be successfully eradicated with cost-effective interventions (Molyneux, 2008).

There is however little evidence which systematically evaluates the impact of treating such diseases. In this paper, I contribute to this debate by studying a historical example of the successful control of a neglected tropical disease, yaws, and examining its impacts on educational attainment in Ghana. A highly infectious disease, yaws was highly prevalent in many parts of the tropics in the first half of the twentieth century. In the Gold Coast colony (now Ghana), it accounted for about 51 percent of reported infectious diseases in hospitals in 1935. Although not fatal, yaws created painful lesions on the skin and joints, and in some sufferers could result in paralysis or permanent disfigurement of the face. The disease is spread through skin contact and was highly prevalent among children aged 5-15 years. Moreover, crowded environments — notably schools and playgrounds — were known to be common sites of yaws transmission (de Vries, 1961; Keeny and Gaan, 1961).

Following World War II, yaws was successfully controlled in many countries in tropcial Africa and Asia, under a global campaign led by the World Health Organization and the United Nations Children’s Fund. The WHO/UNICEF campaign was made possible due to the availability of cheap and mass-produced penicillin which was available in commercial quantities following World War II. Records from the WHO/UNICEF yaws campaign in Ghana indicate that a single dose of intramuscular penicillin was successful in treating overt and latent yaws cases.

This paper examines the long-term impact of successful yaws control in Ghana on eduation outcomes. The identification strategy relies on the cross-district variation in pre-eradication yaws incidence and the observation that penicillin caused an almost immediate reduction in the prevalence of the disease. Following the penicillin campaigns, high yaws prevalence districts, which faced more widespread impairment from yaws, were more likely to benefit from the disease eradication compared with low prevalence regions. Using internal records from the global yaws campaign and microdata from the 2000 Ghana Census, I examine the long-term impact of yaws eradication on educational attainment (specifically, primary school completion, literacy, ability to speak English, and total years of schooling completed). I focus on education outcomes as yaws was most prevalent among children (aged 0-12 years), and schools and playgrounds were known to be a site of yaws disease transmission — partly explaining UNICEF's readiness to support the campaign (de Vries, 1961; Keeny and Gaan, 1961).

In my main results, I observe that a 1 percent decrease in infectious yaws prevalence following the campaigns increased the probability of primary school completion, of being literate, and of being able to speak English by 2.1, 1.9 and 2.1 percentage points respectively. The estimated effect is particularly robust for the female subsample. These results are robust to the inclusion of various time-invariant geographic controls, controls for pre-eradication district characteristics (such as educational attainment, health and agricultural infrastructure) and controls for contemporaneous changes in education and health infrastructure. Using the female subsample, I show that *placebo* campaigns occurring a decade earlier or later than the actual penicillin campaigns do not result in siginificant increases in the educational outcomes. By examining the impacts of the penicillin campaign across different year-of-birth cohorts, I provide further evidence that the marked increases in educational attainment coincide with cohorts born in the early 1950s just prior to the commencement of the penicillin campaigns.

The results of this paper contribute to a recent literature which has examined the long-term impact of past disease eradication programmes, such as hookworm (Bleakley, 2007), and malaria (Bleakley, 2010; Cutler et al, 2010; Lucas, 2010). For hookworm (also, a neglected tropical disease), Bleakley (2007) finds that succesful eradication of this disease in the US south in the early twentieth century increased adult incomes and resulted in improved school enrollment, school attendance, and literacy. For malaria, recent research suggests that malaria eradication increased adult incomes in the Americas (Bleakley, 2010) and also improved consumption for men in India (Cutler et al., 2010). The empirical evidence also suggests that malaria eradication improved educational attainment of women in India (Cutler et al., 2010) and also in Sri Lanka (Lucas, 2010).

The rest of this chapter proceeds as follows: section 3.2 provides a historical review of the yaws penicillin campaign in Ghana. In section 3.3, I discuss the empirical strategy adopted in this paper, and discuss the data used in section 3.4. The main results of this study and robustness checks are presented in sections 3.5 and 3.6 respectively. I provide some discussion of the results in section 3.7 and conclude in section 3.8.

3.2 Yaws and Penicillin Campaign in Ghana

3.2.1 Yaws in Ghana (1900-1950)

Yaws is a disease caused by treponemes or spirochete bacteria called *T. pertenue* which belongs to the family of bacteria that are also causative agents for syphillis and pinta. In the early twentieth century, the disease was endemic in many warm and humid tropical areas of Africa, Central and South America, and Indonesia. Although not fatal, yaws creates lesions in the skin, bone and cartilage and is accompanied by intense pain and itching in the joints. The disease is highly contagious, with the primary mode of transmission being via direct person-to-person contact with the exudate from an infectious yaws lesion. The WHO noted that up to 80 percent of persons exposed to infectious yaws cases contracted the disease and about 10 percent of all yaws cases resulted in permanent disfigurement mostly of the limbs and the face (Perine et al, 1984).

Yaws was highly prevalent in Ghana (Gold Coast colony) in the first half of the twentieth century. The annual reports from the colony's medical department suggest that three diseases were of major concern to the authorities: malaria, yaws and Guinea worm infection. Yaws and Guinea worm infestations caused tremendous morbidity, immobilized communities and threatened food security, especially in the Northern parts of the country (Waddy, 1956). Regarding prevalence, hospital attendance pointed to yaws as the leading cause of morbidity. In 1934, Selwyn-Clarke, the director of the Gold Coast Medical Department noted in his annual report that:

> Yaws continued to hold first pride of place as the condition most frequently met. In point of fact, 60394 cases were treated. This represents about one patient in every four attending for treatment...Malaria still occupied the highest position after yaws on the list of diseases treated at hospitals, dispensaries and welfare centers. (Gold Coast Colony, Report on the Medical Department, 1934, p 9).

Yaws was primarily a disease which afflicted children, with peak incidence occurring the 5-15 year age group. The spread of the disease was also enhanced by crowding and poor sanitation. Upon infection, an initial lesion - called the mother yaw - appeared on the skin after an incubation period of about 21 days. Additional lesions, called secondary yaws, emerge as the bacteria spreads systematically, with each lesions lasting up to 6 months at a time. Overall, the total period of infectiousness of an untreated yaws patient ranged between 12 and 18 months. The disease finally enters an inactive latent phase which can be very destructive: in about 10 percent of cases, it results in crippling and or other permanent disfigurement.

The treatment of yaws in the first half of the twentieth century largely involved the use of intramuscular arsenic injection and oral bismuth preparations. This course of treatment had three major drawbacks: first, the treatments were administered over a three month period, and required repeated vists to the hospitals or dispensaries so there was incomplete adherence to the full course. Second, with arsenic and bismuth preparations side effects such as gingivitis and stomatitis were often reported. Third, and most importantly, these treatments conferred no prophylactic effects so relapses were very common. The treatment was only given to overt yaws cases, neglecting many cases in a latent or incubation state of infection, thus a high level of yaws infection persisted in local communities. As discussed in the next section, all this changed in the 1950s with the successful mass-production of penicillin which proved to be a superior form of treatment, both in efficacy and cost.

3.2.2 WHO/UNICEF Penicillin Campaign (1956-1963)

Penicillin was arguably the most notable medical innovation of the first half of the twentieth century and was able to treat a broad spectrum of bacterial diseases. Although previously isolated by Flemming in 1928, penicillin entered into widespread use only in the 1950s following its succsessful mass production by the US Department of Agriculture during World War II. Following the war, the WHO and UNICEF provided large supplies of penicillin to support a large-scale global anti-yaws campaign in various yaws-endemic countries. In December 1955, at the start of the Global Yaws Campaigns, *Time Magazine* noted that ”the word of penicillin's magic has spread: a single shot, costing only 12 cents, cures a victim of yaws.” (*Africa v. Yaws*, Time Magazine, 26 December, 1955).

In Ghana, the yaws campaign was implemented by teams of mobile public health officials organized as the Medical Field Units (MFU) which had been established in the 1930s by the colonial government. Penicillin used during the campaign was long-acting mixture (procaine penicillin in aluminium monostearate, or PAM) which conferred immunity on asymptomatic household members and supposed latent cases. Compared with prior treatments, penicillin was more efficacious, cheaper and also could be administered very quickly. Besides yaws, a single dose of penicillin would also have effectively treated other bacterial diseases such as pneumonia, meningitis and syphilis (Waddy, 1958). However, the incidence of these diseases was generally low in the Gold Coast, and their concurrent treatment with yaws infections should only introduce very little bias in the analysis below.

For the yaws campaign, the treatment regimen proceeded as follows: villages were given prior notice of the pending MFU visits, and the health officials moved systematically from village to village, examining all residents present. The MFU health officials were comprised of two technical staff: an examiner and an injector. The entire population of a village filed past the examiner, who noted the individual's yaws diagnosis on a piece of paper. An injector further down the queue administered the necessary dosage of penicillin, based on age as well as state of yaws infection[1]. Each village was also re-visited over the next year, and additional penicillin treatments were administered where necessary. The MFU staff documented yaws prevalence rates prior to penicillin treatments, and also during long-term surveys (see Figure 3.1 for map of infection rates). The treatments proved to be very popular and turnouts were often high with long queues of local village residents showing up for treatments early in the morning[2].

3.2.3 Related Studies

There has been recent interest in examining the economic burden and long-term consequences of tropical diseases. Much of the recent work has focused on two diseases: malaria (Bleakley, 2010; Lucas, 2010; Cutler et al, 2010) and helminthic infections (Bleakley, 2007; Miguel and Kremer, 2004). The combined evidence on malaria suggests that

[1] The most common method recommended by the WHO was the Total Mass Treatment regimen, in which adult yaws cases received $4cm^3$ of penicillin and children received $2cm^3$. Non-yaws cases received half of this dosage (Scott, 1959).

[2] Yaws was never completely eradicated in Ghana, and there was a brief resurgence in isolated villages in the early 1980s (Agadzi et al, 1985). The overall incidence in affected regions was however always below the WHO threshold of 0.5 percent.

malaria eradication improved adult incomes in the Americas (Bleakley, 2010) and also improved consumption for men in India (Cutler et al., 2010). On education outcomes, existing evidence suggests that malaria eradication improved educational attainment of women in India (Cutler et al., 2010) and also in Sri Lanka (Lucas, 2010). For helminthic infections (also considered neglected tropical diseases), Bleakley (2007) estimates that hookworm eradication in the American South resulted in increases in various education outcomes and increased incomes. Miguel and Kremer (2004) similarly find positive treatment externalities in treating parasitic worm infections in Kenya, and this treatment also improved school attendance by children.

More broadly, there has been recent policy interest in work on neglected tropical diseases which encompass all major tropical diseases apart from the three diseases frequently funded by donor agencies, namely: tuberculosis, HIV/AIDS and malaria[3]. Public health analysts have argued about the potential socioeconomic benefits of tackling such diseases which afflict the majority of the world's poor (Hotez et al., 2009). However there is only limited work which examines the economic consequences of the control of any such diseases. In this context, the largely successful yaws penicillin campaign in Ghana provides a useful historical case study.

3.3 Empirical Strategy

The identification strategy adopted here exploits the cross-district differences in yaws infection prior to the WHO/UNICEF penicillin campaign to examine the long term effect of yaws control. I use a differences-in-differences model to compare human capital attainment of various year-of-birth cohorts from various districts of Ghana. Since the penicillin campaign effectively reduced infectious yaws rates to less than 0.01 percent in all districts, areas with high initial yaws prevalence were likely to experience greater benefits from yaws eradication compared with low prevalence areas. Moreover, the introduction of the peni-

[3]Specifically, the World Health Organization classifies the following as neglected tropical diseases: blinding trachoma, Buruli ulcer, Chagas disease, dengue, dracunculiasis, human African trypanosomiasis, Japanese encephalities, leishmaniases, leprosy, lymphatic filiarisis, onchocerciasis, schistosomiases, soil-transmitted helminthiasis and yaws (Liese et al, 2010)

cillin treatment by an international agency eliminates potential endogeneity concerns which may have arisen, for example, if yaws eradication was triggered by increases in incomes of farmers of cash crop-growing regions. Following the recent literature on the economic impact of tropical disease eradication, this study also focuses on childhood exposure since the medical literature noted yaws as a disease which predominantly afflicted children and could potentially discourage school attendance given its contagiousness and its ability to inflict permanent deformities.

I specifically compare outcomes at a given point in time for various year-of-birth cohorts from districts with distinct pre-campaign yaws infection levels. I construct a binary treatment variable which equals one for cohorts born after the penicillin campaigns[4]. Although yaws infection data is available at the village level, pre-campaign yaws infection is aggregated to the district level since our outcomes data only provide us with each individual's district of residence (see section 3.4 for further discussion). Thus, an identifying assumption in this study is that an individual's district of residence is highly correlated with their district of birth and education.

I omit cohorts born during the active years of the campaigns (1955-65), and restrict my analysis to the sample of individuals born a decade prior to, and a decade following, the penicillin campaigns. For a given individual i, in district, d, and year-of-birth cohort, c, I run a regression of the following form:

$$Y_{idc} = \beta(Post_{dc} \times Inf_d) + \delta_d + \delta_c + ((Post_{dc} \times X_d^{'})\lambda + \epsilon_{idc} \tag{3.1}$$

where the dummy $Post_{dc} = 1$ if year of birth occurred after the commencement of the penicillin campaign in district d for members of cohort, c; Inf_d refers to pre-campaign yaws prevalence rate in district d; and δ_d and δ_c are respectively district and birth cohort fixed effects. X_d is a vector containing various district-level controls such as time-invariant

[4]Previous work which utilize a similar difference-in-difference estimation strategy to examine disease eradication impacts adopt a dose response model to study the precise effect of exposure to the disease eradication campaigns, see for example, Bleakley (2010). Such an approach requires precise data on the implementation dates for the treatment in various districts in the country. The majority of the penicillin campaigns were implemented in Ghana over the period 1956-1960, however, the timing of the program roll out across various districts in the country are not exactly known (see Fasquelle, 1973) thus making it difficult to implement an approach based on precise dose of exposure.

geographic controls, and distance from the centroid of each cocoa district to the nearest agricultural station, railway station and hospital, as at 1948. The geographic controls include latitude, longitude, and mean area-weighted elevation and slope. The coefficient of interest is β which provides the difference-in-difference estimate of the impact of the penicillin campaign on the outcome variable. I report results in equation (3.1) above as my baseline specification, and further run it separately for various demographic subgroups (namely: males, females, the Akan ethnic group, and Christians). In robustness checks, I control for several other pre-treatment characteristics (which enter into equation (3.1) interacted with the dummy for $Post_{dc}$ campaign), for contemporaneous changes in health and education infrastructure, and for possible mean reversion.

Next, I extend the analysis to examine a more flexible specification which allows the impact of the peniclin campaigns to vary for different age cohorts. Using the female subsample, I examine cohorts aged between 15 and 90 grouped into 14 age categories at five-year intervals, with persons aged above 80 serving as the omitted category. I run the following regression:

$$Y_{idc} = \beta_g \sum_{g=1}^{13} (Post_{dg} \times Inf_d) + \delta_d + \delta_c + ((Post_{dg} \times X_d^{'})\lambda + \epsilon_{idc} \qquad (3.2)$$

where the dummy $Post_{dg} = 1$ if an individual in district, d, was born within the five-year age interval, g; Inf_d refers to pre-campaign yaws prevalence rate in district d; and δ_d and δ_c are respectively district and birth cohort fixed effects. X_d is a vector containing various district-level controls such as time-invariant geographic controls.

With this specification, I am able to examine how the relationship between yaws infection rates and the outcome variable changed for cohorts born in various years prior to and following the penicillin campaigns. In this case, I interact the control variables with a dummy for each age category cohort, so that the effects of the controls are allowed to vary across different cohorts. For each cohort, the coefficient β_c provides the impact of the yaws eradication. Thus, we expect no impact of the treatment for cohorts born prior to 1950 when the penicillin campaigns begun, and for the coefficient to be positive and significant in subsequent periods. As the penicillin campaigns were implemented across various districts in Ghana over the period 1956 to 1963, we would expect the trend in β_c

to change for cohorts born around 1950, who were most likely to be exposed to penicillin during the age period of primary school attendance.

How plausible is the assumption that the risk of yaws infection served to reduce school enrollments and thus lower human capital investments in childhood? This assumption appears credible as yaws was predominantly a disease contracted by children, and was known to be highly contagious. Moreover, it was also known that crowded environments — notably schools — provided an ideal setting for the disease transmission (Soetopo and Wasito, 1953; De Vries, 1961). Schools, supported by the colonial government as well as missionary organizations, had been established in most towns in Ghana by the 1960s, even in the relatively poorer northern regions (Scott, 1959). The debilitating effects of yaws in the Gold Coast were noted by Saunders (as cited in Hackett (1953, pp 150)):

> Dr. G. F. T. Saunders regards the suffereing from yaws as very great and includes rheumatic pains in children, and chronic "rheumatism", ulcers, and deformities in adults. He regards yaws as a constant drain on the economic resources of a community. In one area, after mass treatment, there was a reduction in tertiary yaws, and it was said that the chief had less difficulty in getting work done at periods of intensive farm-work; the population attributed this to the effects of the anti-yaws campaign.

The popularity of the penicillin campaign and the nearly perfect turnout rates encountered during treatment further suggests that many households were eager to avoid the long-term debilitating effects of yaws infection.

A main identification assumption for this study is that apart from the WHO/UNICEF campaign, there were no differential changes in human capital investments across localities which were correlated with existing yaws infection rates. This assumption needs to be addressed since the decade 1950-1960 also marked a period of rapid expansion in public goods infrastructure in the newly independent Ghanaian state. Areas with higher yaws infections tended to be poorer and also in the northern parts of the country (see Table 1) where standards of hygiene were generally lower due to inadequate water supplies. Any concurrent expansion of public services in these regions during the campaigns may therefore lead to an upward bias in the estimates of the penicillin treatment effect. I therefore utilize controls for contemporaneous health and infrastructure investments in the 1950s and 1960s in additional robustness checks later in this chapter.

3.4 Data

3.4.1 Yaws Infection Rates

Data on yaws infection prevalence is obtained from published annual reports of the Gold Coast/Ghana Medical Field Units (MFUs) and from internal publications of WHO epidemiologists. The MFU annual reports from 1959-1964 provided aggregate data for yaws infections throughout the country. Data from the annual reports were complemented by epidemiological data summarized by physicians working on the yaws campaign (Scott (1959), Christiansen (1963), Rosei (1963) and Fasquelle (1973)). The internal epidemiological reports provides details of pre- and post-eradication yaws infection rates at the village/town level and aggregated to the district and regional levels. The yaws campaign proceeded with an initial treatment survey in which all members of a given community were initially examined, with various cases of yaws diagnosed and treated. The initial survey was followed by two surveillance surveys and concluded with a long term survey after about a year to track any outstanding latent or reservoir infections. Data on yaws prevalence was presented in two forms: either as the prevalance rate of *infectious* yaws or as the prevalence of *total* yaws which comprised of infectious and non-infectious cases (such as late-stage hyperkeratosis).

Figure 3.1 and Table 3.1 provide a summary of yaws infection rates across all districts and regions in Ghana circa 1960. The mean (simple average) prevalence rates of total yaws and infectious yaws across all districts were 8.9 and 0.7 percent respectively. The disease prevalence was higher in northern districts which tended to be poorer and more distant from existing transport infrastructure. Since yaws was predominantly a disease affecting children, and prevalence rates are calculated for the entire population, the incidence of the disease among children was probably nearly twice as high as the recorded prevalence rates. The regressions conducted in this study use data on infectious yaws (rather than total yaws) which were consistently provided for all districts during the penicillin campaign. Table 3.1 presents summary statistics.

TABLE 3.1: SUMMARY STATISTICS

VARIABLES	All Districts (1)	Low Districts (2)	High Districts (3)	*P*-value (2) vs (3) (4)
Yaws infection rates				
Infectious yaws prevalence rate (in percent)	0.721	0.326	1.129	0.0000
Geographic characteristics				
Longitude	-0.992	-1.185	-0.795	0.0878
Latitude	6.936	6.602	7.280	0.1030
Slope (in degrees)	2.899	2.753	3.049	0.4025
Elevation (in meters)	158.6	127.2	190.9	0.0012
Landarea (sq. km.)	1,971	1,404	2,554	0.0352
Cocoa suitability index	37.02	31.94	42.25	0.2476
Fraction of land as forest reserve	0.071	0.072	0.070	0.9551
Proximity to infrastructure				
Distance to railway (km)	113,227	109,699	116,858	0.8376
Distance to road (km)	49,884	55,683	43,915	0.6171
Distance to mission station (km)	68,666	73,554	63,634	0.5494
Distance to cocoa station (km)	102,978	113,362	92,287	0.4420
Distance to agricultural station (km)	29,966	29,491	30,456	0.8237
Distance to Accra (km)	224,989	222,042	228,024	0.8833
Economic characteristics				
Fraction of pop. employed in cocoa, (1948 census)	0.127	0.109	0.145	0.1530
Fraction of pop. employed in farming, (1948 census)	0.585	0.595	0.575	0.6696
Fraction of pop. employed in mining, (1948 census)	0.0551	0.0692	0.0406	0.4469
Fraction of pop. employed in manufacturing	0.0580	0.0591	0.0569	0.7750
Fraction of pop. employed in services	0.1034	0.1072	0.0996	0.6316
Acreage under cocoa cultivation (sq. km.)	190.9	154.4	228.6	0.1506
Local government revenue (1948, Sh/capita)	0.460	0.415	0.506	0.3960
Demographic characteristics				
Fraction of Christians (1940 estimates)	0.5913	0.6218	0.5608	0.3378
Fraction of Moslems (1940 estimates)	0.1571	0.1109	0.2032	0.0410
Fraction of Akan ethnic group (1940 estimates)	0.4062	0.4591	0.3532	0.1569
Primary school completion (1940 estimates)	0.185	0.199	0.169	0.2730
Literacy rate (1940 estimates)	0.264	0.274	0.254	0.5580
Able to speak english, in percent (1940 estimates)	0.224	0.2380	0.2099	0.3684
Years of schooling completed (1940 estimates)	2.354	2.5432	2.1650	0.2521
Observations	69	35	34	

NOTES: This table provides descriptive statistics for pre-treatment characteristics. Data on yaws infection rates are compiled from Scott (1959) for the Northern territories, Onori (1962) for the Volta region, and Rosei (1963) for the Brong Ahafo region, and from Fasquelle (1973) and Christiansen (1963). Data on geographic characteristics obtained by overlaying administrative map of Ghana/Gold Coast on elevation (slope) data obtained from SRTM, and computing mean area-weighted elevation (slope). Other demographic characteristics obtained from the 1948 Gold Coast Census, and author's calculations based on IPUMS microsample for the 2000 Ghana census.

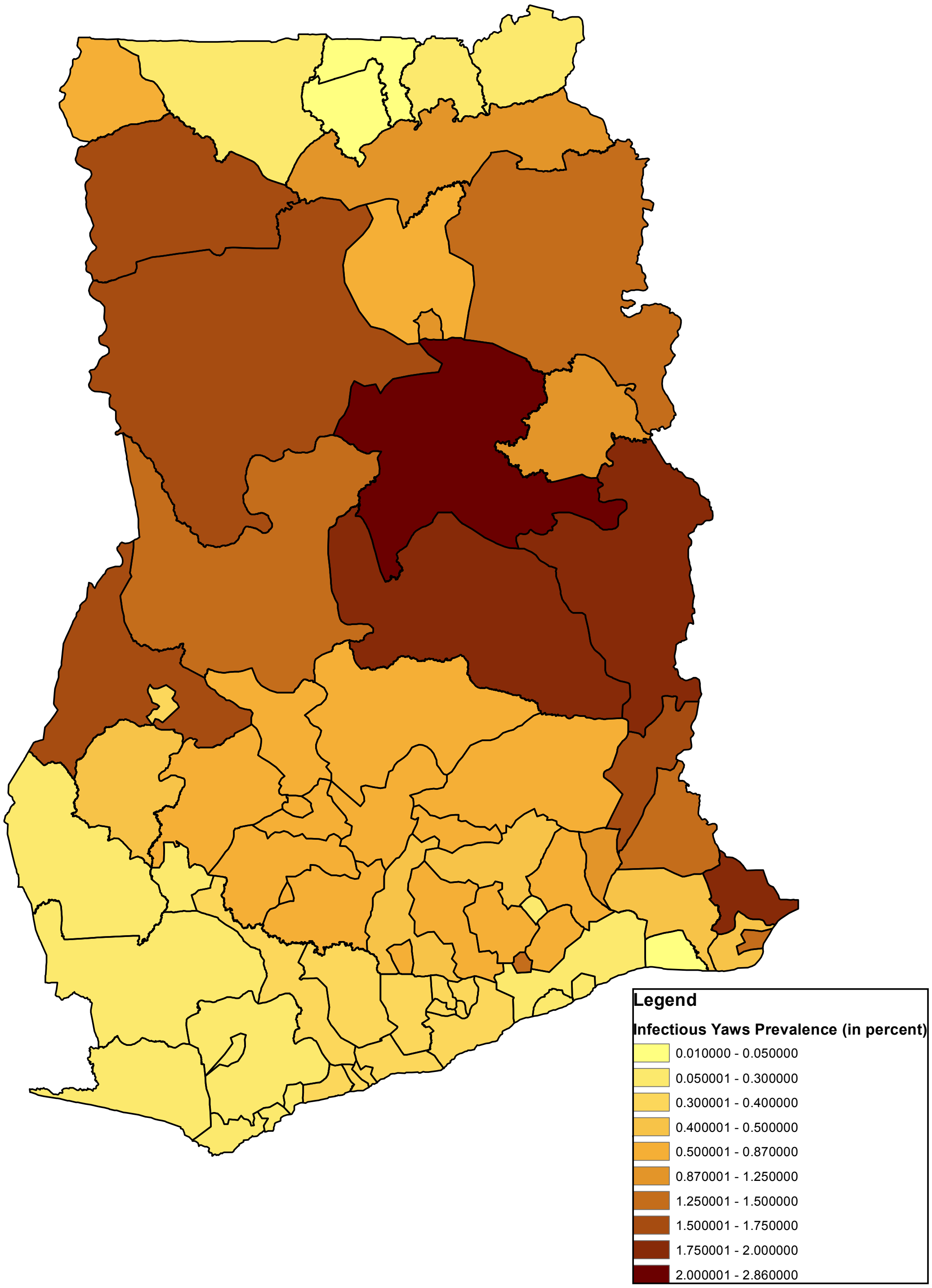

Figure 3.1: **Yaws Prevalence Rate in Ghana, 1956.** Data on yaws infection rates are compiled from Scott (1959) for the Northern territories, Onori (1962) for the Volta region, and Rosei (1963) for the Brong Ahafo region, and from Fasquelle (1973) and Christiansen (1963).

3.4.2 Outcomes

For outcome measures, I utilize data from the IPUMS microsample of the 2000 Ghana census. The IPUMS microsample of the 2000 Ghana Census provides data on various measures of human capital attainment (primary school completion, literacy, ability to speak English, and years of total schooling), and comprises observations on about 1.3 million persons.

3.4.3 District Pre-treatment characteristics

Data on pre-treatment district population and economic characteristics are obtained from the *Gold Coast Census of 1948*. Data on various time-invariant geographic characteristics are obtained by overlaying the 1948 administrative map of Ghana/Gold Coast on elevation and slope data obtained from SRTM/ArcMaps and computing the mean area-weighted elevation and slope, as well as the longitude and latitude of the centroid of each district. I also calculate various district-level geographic controls, including the distance from the centroid of each district to the nearest railway line, to the nearest protestant mission, to the nearest agricultural station and to the nearest hospital in 1948. The 1948 *Gold Coast Census* also provides data on pre-treatment economic activity such as the fraction of the population employed in cocoa, in agriculture and in mining. I compute the suitability index of each district for cocoa cultivation using the cocoa suitability index developed by the Ghana Soil Research Institute. The index provides a 0-100 score of the suitability for cocoa farming on a GIS grid raster, taking into account local precipitation, temperature and soil quality. I also calculate various demographic characteristics such as religious and ethnic compositions, and educational attainment, using data from the IPUMS microsample of the 2000 Ghana census.

3.4.4 Education and health infrastructure investments

Data on number of schools, classrooms, teachers are obtained from publications by the Ministry of Education for the years 1962, 1968, and 1972. Education districts differ from administrative districts used in census, and so I aggregate results to the regional level.

To calculate teachers per capita and classrooms per capita, I normalize by children in the relevant age category. School density is the number of schools divided by land area of administrative region (in square kilometers).

Health infrastructure data is obtained from 2 sources. For 1951, I use government findings as published in the Report of the Commission of Enquiry into the Health Needs of the Gold Coast. For 1963, I obtain data from Pawel (1963). I calculate hospital density (number of hospitals divided by land area of administrative region), and per capita number of hospital beds. For both 1951 and 1963 health data are for general hospitals (comprised of government hospitals, mission hospitals and hospitals run by mining companies). Data on cocoa production in 1951 and 1963 for various cocoa-producing regions are obtained from the *Analysis of Cocoa Purchases by Societies, Districts and Regions for 1947-1965* published by the Ghana Cocoa Board.

3.5 Results

3.5.1 Difference-in-Differences Analysis

Results of the difference-in-difference specification in equation (3.1) are presented in Table 3.2. I report results for the baseline specification in Panel A, while Panel B presents results separately for various demographic subgroups. The baseline results include controls only for time-invariant district geographical characteristics (longitude, latitude, slope and land area). These controls enter equation (3.1) interacted with the dummy for $Post_{dc}$. In each panel, I present results for the probability of primary school completion, for literacy rates, ability to speak English, and total years of schooling completed. The results based on the specification in equation (3.1) indicate that yaws treatment improved various educational outcomes, with the baseline results significant at the 5 percent level in columns (1)-(4). The reported point estimates are for the coefficient, β, on the $(Post_{dc} \times Inf_d)$ term, which provides the impact of being born a decade following the penicillin campaigns versus a decade prior to the campaigns in a high endemicity district, compared to the effect in a low endemicity district. A one percentage point increase in initial infectious yaws incidence was associated with a post-campaign increase in the probability of primary school completion of 2.1 percentage points. Literacy and the ability to speak English similarly increased by 1.9 and 2.1 percentage points. In the baseline regression, I also find that a one percentage point increase in yaws prevalence prior to the campaigns resulted in an increase of 0.22 years of schooling. Prior to the penicillin campaigns, the median infectious yaws incidence rate was 0.78 percent. Thus, for the hypothetical district with the median level of yaws incidence, mean schooling attainment improved by 0.17 years following the penicillin campaigns. Results for the demographic subgroups are reported in Panel B. The baseline results appear to be driven by the impact of the penicillin campaigns on females. For the case of girls, I find that the probability of primary school completion, of being literate and being able to speak English increased by 2.1, 2.3 and 2.1 percentage points respectively. In addition, a 1 percentage point increase in initial infectious yaws prevalence resulted in an increase of total years of schooling completed by 0.2 years for girls following the penicillin campaigns.

3.6 Robustness Checks

I conduct a number of robustness checks to address several sources of bias in the estimates reported above. I show that the results obtained in the previous section are not sensitive to controlling for a number of pre-treatment district controls, for restricting the sample to the population of non-movers, as well as to controlling for mean reversion.

3.6.1 Sensitivity Analyses

In this section, I present results which introduce additional controls into the baseline specification reported above. This is important for four reasons. First, the pre-treatment levels of yaws infections are geographically clustered in northern districts of the country. The pre-existing disease intensity levels may therefore be correlated with the expansion of economic activity and missionary influences which occurred in Ghana in the first half of the twentieth century but mostly in southern districts. Thus, the first set of robustness exercises report results controlling for various pre-treatment district characteristics such as proximity to the existing railway network, to mission stations, to agricultural stations and

TABLE 3.2: PENICILLIN EXPOSURE AND HUMAN CAPITAL ATTAINMENT

VARIABLES	(1) Primary school completion	(2) Literacy	(3) Able to speak English	(4) Years of schooling
		PANEL A: MAIN RESULTS		
Baseline	0.0214***	0.0193***	0.0205***	0.223***
	(0.00696)	(0.00621)	(0.00642)	(0.0797)
Observations	447,408	447,408	447,408	447,408
R-squared	0.185	0.212	0.198	0.168
District fixed effects	YES	YES	YES	YES
Year of birth fixed effects	YES	YES	YES	YES
Geographical controls	YES	YES	YES	YES
		PANEL B: DEMOGRAPHIC SUBGROUPS		
Males	0.0174*	0.0108	0.0153**	0.200*
	(0.00914)	(0.00659)	(0.00732)	(0.115)
Females	0.0210***	0.0231***	0.0212***	0.201**
	(0.00688)	(0.00704)	(0.00739)	(0.0785)
Akan	0.0343***	0.0331***	0.0403***	0.473***
	(0.0116)	(0.0115)	(0.0126)	(0.139)
Christian	0.0292***	0.0243***	0.0286***	0.290***
	(0.00826)	(0.00796)	(0.00812)	(0.0982)
District fixed effects	YES	YES	YES	YES
Year of birth fixed effects	YES	YES	YES	YES
Geographical controls	YES	YES	YES	YES

NOTES: Data is obtained from the IPUMS microsample of the 2000 Ghana census. Results are based on the regression model in equation (3.1). The coefficient on $(Post_{dc} \times Inf_d)$ provides the difference-in-difference estimate comparing the mean outcome variable for younger cohorts (born a decade following the penicillin campaigns) to an older cohort (born a decade prior to the penicillin campaigns) in a high endemicity district, compared to the difference in a low endemicity district. Geographic controls are for latitude, longitude, land area and mean area-weighted slope. Clustered standard errors, using the 2000 district administrative boundaries, are in parentheses. *** p<0.01, ** p<0.05, * p<0.1.

to hospitals as well as controls for district cocoa soil suitability.

Second, the 1950s also marked a period of rapid expansion in public infrastructure in the Ghana/Gold Coast colony. If such investments were primarily targeted at localities with high yaws infection rates, then such programs may confound the estimates obtained above, and lead to an overestimation of the human capital gains attributable to the penicillin campaigns. I therefore repeat the regressions using specifications in equation (1) but now with region-of-birth controls for changes in health and education infrastructure. The health controls are for changes in hospital density (i.e. number of hospitals divided by land area of the administrative region) and per capita number of hospital beds. The education controls are for changes in school density, teachers per capita and schools per capita.

Third, legislative elections were also introduced in the Gold Coast colony in the 1950s, resulting in the development of local political movements (Austin, 1964). The results in this paper could also be biased if areas with higher yaws infection, which were poorer, tended to support the nationalist movement of Nkrumah's Conventions People's Party (CPP) and therefore received larger targeted transfers from the central government. I therefore report results from additional specifications which control for the CPP vote share in the 1956 Gold Coast Legislative Elections.

Finally, I also explore the possibility of expansion in cocoa production as a possible explanation for our results. Although cocoa production had commenced in the Gold Coast colony at the start of the twentieth century, there were substantial expansions in cultivated acreages in the 1960s. The frontier for cocoa production moved westwards as new forests in the Western and Brong Ahafo regions were cleared for cultivation. Other cocoa-growing areas (particularly in the Eastern Region) experienced a temporary decline in output as a result of the swollen shoot disease which attacked cocoa trees (Ross and Broatch, 1951; Tanburn, 1957). The logarithmic change of cocoa output (for the period 1953-1961) is therefore included as an additional control to the base specification. All controls enter equation (1) interacted with a dummy for post-campaign.

In Table 3.3 Panels A-D, column (1) repeats the previous baseline results controlling only for time-invariant geographical characteristics[5]. Consider Panel A on primary school

[5]These controls enter equation (3.1) interacted with the dummy for $Post_{dc}$.

completion. In column (2), when controls for cocoa soil suitability are introduced, the point estimate remains highly significant, although slightly reduced in magnitude. Next, in columns (3) and (4), to account for the relative remoteness of various districts, I control for proximity to the rail transport infrastructure (as at 1948) and for distance to Accra. The point estimates remain robust to the inclusion of these additional controls. In columns (5) and (6), I further control for proximity to mission stations, hospitals and agricultural stations, which may serve as a proxy for pre-eradication availability of various social services. Column (7) includes all controls above. Comparing columns (1) and (7) for Panels A-D, the results remain robust and significant in both columns, although the magnitude of the point estimates are slightly reduced in the case of primary school completion and years of schooling attained.

Table 3.4 repeats the baseline specification in equation (1) but with controls for contemporaneous changes in health and education infrastructure (over the period 1955-1965), for changes in cocoa output, as well as for the CPP vote share in the 1956 Elections. Column (1) of Table 3.4 repeats the regression specification from column (7) of Table 3.3. Next, in columns (2) to (4), I successively control for changes in the number of schools, number of teachers, and number of hospitals. Columns (5) and (6) control for changes in cocoa output and the CPP vote share respectively. For Panels (A) to (D), the results remain robust and significant when viewed across from columns (1) to (6), and thus provide further support for the impact of the penicillin campaign in improving educational attainment.

TABLE 3.3: ROBUSTNESS CHECKS

	(1)	(2)	(3)	(4)	(5)	(6)	(7)
PANEL A: Primary school completion							
$Post_{dc} \times Inf_d$	0.0214***	0.0181***	0.0184***	0.0222***	0.0192***	0.0228***	0.0165**
	(0.00696)	(0.00650)	(0.00683)	(0.00702)	(0.00642)	(0.00681)	(0.00643)
PANEL B: Literacy							
$Post_{dc} \times Inf_d$	0.0193***	0.0179***	0.0188***	0.0195***	0.0183***	0.0202***	0.0191***
	(0.00621)	(0.00615)	(0.00634)	(0.00621)	(0.00617)	(0.00598)	(0.00603)
PANEL C: Ability to speak English							
$Post_{dc} \times Inf_d$	0.0205***	0.0191***	0.0196***	0.0210***	0.0195***	0.0215***	0.0204***
	(0.00642)	(0.00643)	(0.00647)	(0.00637)	(0.00638)	(0.00625)	(0.00640)
PANEL D: Years of schooling							
$Post_{dc} \times Inf_d$	0.223***	0.199**	0.211**	0.222***	0.207***	0.231***	0.182**
	(0.0797)	(0.0764)	(0.0809)	(0.0793)	(0.0788)	(0.0782)	(0.0764)
District fixed effects	YES	YES	YES	YES	YES	YES	YES
Year of birth fixed effects	YES	YES	YES	YES	YES	YES	YES
Geographical controls	YES	YES	YES	YES	YES	YES	YES
Cocoa soil suitability	NO	YES	NO	NO	NO	NO	YES
Distance to rail	NO	NO	YES	NO	NO	NO	YES
Distance to Accra	NO	NO	NO	YES	NO	NO	YES
Distance to mission station/hospital	NO	NO	NO	NO	YES	NO	YES
Distance to agricultural station	NO	NO	NO	NO	NO	YES	YES

NOTES: Data is obtained from the IPUMS microsample of the 2000 Ghana census. Results are based on the regression model in equation (3.1). The coefficient on $(Post_{dc} \times Inf_d)$ provides the difference-in-difference estimate comparing the mean outcome variable for younger cohorts (born a decade following the penicillin campaigns) to an older cohort (born a decade prior to the penicillin campaigns) in a high endemicity district, compared to the difference in a low endemicity district. Geographic controls are for latitude, longitude, land area and mean area-weighted slope. Cocoa soil suitability controls is based on an index computed by the Ghana Soil Research Institute, which provides a measure for the overall suitability for cocoa farming. Additional controls are for the distance from the centroid of each district to the nearest railway line, to Accra, to the nearest mission station/hospital, and to the nearest agricultural station. Robust standard errors are in parentheses, and clustered using district administrative boundaries (as at 2000). *** p<0.01, ** p<0.05, * p<0.1.

3.6.2 Non movers

I also conduct additional robustness checks where I repeat the baseline specification but omitting all migrants, and examine the results for the nonmover sample. The estimates of the effect of the penicillin campaign would be biased upwards if selective migration occurred with healthier or more capable persons migrating to previously endemic regions following the campaign. Using the IPUMS data, I restrict the sample to persons who live in their same district of birth. Results for this nonmover sample is presented in Table 3.5, Panel A. The point estimates of the effect of the penicillin campaigns remain positive and significant for the various education outcomes examined.

3.6.3 Mean Reversion

Panel B of Table 3.5 present results which correct for possible mean reversion in educational attainment. Mean reversion may be a concern if for example older cohorts in high endemic areas experienced a temporary shock which resulted in higher yaws infections, reduced schooling and low productivity. In this case, increases in educational attainment would have been expected from highly infected areas even in the absence of the penicillin campaign. To correct for possible mean reversion, I include the pre-treatment level (as at 1940) of the outcome variable as controls (interacted with the dummy for $Post_{dc}$). Thus, in Panel B of Table 5, I report results which control for the pre-1940 levels of primary school completion, literacy, ability to speak English and years of schooling in columns (1) to (4) respectively. The point estimates controlling for mean reversion are somewhat smaller in magnitude, but remain significant as in my main results in Table 3.3.

3.6.4 Placebo Eradication Campaigns

Using the female subsample, I examine the impact of two placebo eradication campaigns which may have occurred a decade prior, or decade following, the actual penicillin campaigns. I focus on two outcome variables — the probability of primary school completion and the total years of schooling completed. Panels (A), (B) and (C) respectively report difference-in-difference results for the actual penicillin campaigm (1955-1965), for

TABLE 3.4: CONTROL FOR CONTEMPORANEOUS CHANGES

	(1)	(2)	(3)	(4)	(5)	(6)
PANEL A: Primary school completion						
$Post_{dc} \times Inf_d$	0.0165**	0.0157**	0.0138**	0.0152**	0.0166***	0.0168**
	(0.00643)	(0.00637)	(0.00685)	(0.00633)	(0.00633)	(0.00644)
PANEL B: Literacy						
$Post_{dc} \times Inf_d$	0.0191***	0.0196***	0.0161**	0.0198***	0.0191***	0.0205***
	(0.00603)	(0.00612)	(0.00617)	(0.00627)	(0.00604)	(0.00599)
PANEL C: Ability to speak English						
$Post_{dc} \times Inf_d$	0.0204***	0.0206***	0.0171**	0.0207***	0.0204***	0.0222***
	(0.00640)	(0.00647)	(0.00658)	(0.00658)	(0.00641)	(0.00639)
PANEL D: Years of schooling						
$Post_{dc} \times Inf_d$	0.182**	0.171**	0.134*	0.170**	0.184**	0.201***
	(0.0764)	(0.0740)	(0.0786)	(0.0750)	(0.0749)	(0.0759)
District fixed effects	YES	YES	YES	YES	YES	YES
Year of birth fixed effects	YES	YES	YES	YES	YES	YES
Controls from Table 3, Col. (7)	YES	YES	YES	YES	YES	YES
Change in number of schools	NO	YES	NO	NO	NO	NO
Change in number of teachers	NO	NO	YES	NO	NO	NO
Change in number of hospitals	NO	NO	NO	YES	NO	NO
Change in cocoa output	NO	NO	NO	NO	YES	NO
Vote share for CPP Party	NO	NO	NO	NO	NO	YES

NOTES: Data is obtained from the IPUMS microsample of the 2000 Ghana census. Results are based on the regression model in equation (3.1), and include controls for contemporaneous changes in education and health infrastructure, and for political support for Nkrumah's CPP party. The coefficient on $(Post_{dc} \times Inf_d)$ provides the difference-in-difference estimate comparing the mean outcome variable for younger cohorts (born a decade following the penicillin campaigns) to an older cohort (born a decade prior to the penicillin campaigns) in a high endemicity district, compared to the difference in a low endemicity district. Robust standard errors are in parentheses, and clustered using district administrative boundaries (as at 2000). *** p<0.01, ** p<0.05, * p<0.1.

an early placebo campaign (1945-1955) and for a later placebo campaign (1965-1975). As before, the point estimates for the actual campaign are statistically significant (as in Table 3.5, controlling for mean reversion). In both placebo campaigns, the coefficient β, on the $(Post_c \times Inf_d)$ term is much smaller in magnitude and never statistically significant.

3.6.5 Cohort Analysis

In this section, I examine how the impact of the penicillin campaign on human capital attainment varied across year of birth cohorts. As shown in equation (3.2) in section 3.3, I use the IPUMS Ghana 2000 Census microsample to examine cohorts aged between 15 and 90 grouped into 14 age categories at five-year intervals. The omitted category refers to individuals aged above 80. For each five-year cohort, the coefficient β_c of $(Post_{dc} \times Inf_d)$ provides the difference-in-difference estimate of yaws eradication.

We would expect no impact of the treatment for cohorts born prior to 1950 when the penicillin campaigns begun, and for the coefficient to be positive and significant in subsequent periods. Specifically, as the penicillin campaigns were implemented across various districts in Ghana over the period 1956 to 1963, we would expect the trend in β_c to change for cohorts born around 1950, who were most likely to be exposed to penicillin during the age period of primary school attendance.

Figures 3.2 and 3.3 plots the coefficients β_c for female cohorts born between 1920 and 1985, for two outcome variables: the probability of primary school completion and the total years of schooling completed. Each data point represents the cohort specific β_c while the x-axis refers to the year of birth (in five-year bins). The rise in the magnitude of the coefficients begins for cohorts born between 1935-1940, however, the coefficients β_c are small and insignificant for cohorts born prior to 1950. In Figure 3.2, the coefficient becomes statistically significant for cohorts born precisely in the early 1950s who were most likely exposed to the penicillin campaigns during the period of primary school enrollment.

Notice that the results reported in Figures 3.2 and 3.3 include various controls (for district pre-treatment characteristics, for contemporaneous changes in cocoa output, health and education infrastructure, and also for the CPP vote share from the 1956 Legislative Elections), and also control for possible mean reversion as shown in Table 3.4. Thus concern that the results might reflect pre-existing trends or contemporaneous public infrastructure investments is partly ameliorated by the robustness of the difference-in-difference effects to the extensive set of controls.

3.7 Discussion

Taken together, the results obtained in this study suggest that the WHO/UNICEF yaws campaign improved education outcomes of cohorts born following the penicillin treatments. To provide some broader context, the WHO/UNICEF yaws campaign was also conducted in several other countries in other endemic countries Africa, Asia and the Caribbean - many of which had much higher yaws disease burdens than reported in the Ghana. For example, in the Nsukka region of Eastern Nigeria, Zahra (1956) reported a mean incidence rate of infectious yaws of about 2.5 percent and total yaws infection rates of nearly 20 percent. For such high prevalence rates as recorded in Nsukka, the results from this study suggest that yaws eradication improved the probability of primary school completion and total years of education by about 5.3 percentage points and 0.5 years respectively. The magnitude of these increases are comparable with the estimates from the possible education benefits from control of say, malaria. For example, Lucas (2010) estimates that complete eradication of malaria in Uganda (which recorded the highest malaria prevalence of 47.8 percent in 2005) would result in individual gains of about 0.5 years of additional schooling.

It is difficult to extend the results above on yaws control in Ghana to the possible benefits that may arise from the eradication of other neglected tropical diseases. However, a common feature of yaws eradication in the mid-twentieth century and some neglected tropical diseases today (such as onchocerciasis and filariasis) is that they are both highly cost-effective interventions, which can be readily administered to entire communities or integrated into existing health campaigns. For example, for the Global Yaws Campaign, UNICEF estimated per capita treatment costs of about \$0.67 in 1958 dollars or about \$ 4.60 in 2008 dollars (Keeny and Gaan, 1961). These costs exceed existing cost estimates of about \$1.26 (in 2006) per treatment with invermectin for the WHO onchocerciasis control program in Africa (Conteh et al., 2010).

TABLE 3.5: NON MOVER SAMPLE AND CONTROLS FOR MEAN REVERSION

	(1) Primary school completion	(2) Literacy	(3) Able to speak English	(4) Years of schooling
Panel A: Non-mover sample				
$Post_{dc} \times Inf_d$	0.0184***	0.0220***	0.0237***	0.210***
	(0.00642)	(0.00595)	(0.00647)	(0.0745)
Observations	414,430	414,430	414,430	414,430
R-squared	0.187	0.215	0.202	0.170
Panel B: Controls for Mean Reversion				
$Post_{dc} \times Inf_d$	0.0130*	0.0154**	0.0165**	0.128*
	(0.00665)	(0.00594)	(0.00644)	(0.0746)
Observations	447,408	447,408	447,408	447,408
R-squared	0.185	0.212	0.198	0.168
District fixed effects	YES	YES	YES	YES
Year of birth fixed effects	YES	YES	YES	YES
Geographical controls	YES	YES	YES	YES
Controls for cocoa soil suitability	YES	YES	YES	YES
Controls for dist. to Accra	YES	YES	YES	YES
Controls for dist. to missions/hospitals	YES	YES	YES	YES
Controls for dist. to agricultural station	YES	YES	YES	YES

NOTES: Results in Panel A are for the non-mover sample, repeating the specification in equation (3.1) with all controls but omitting all migrants. Panel B reports results controlling for mean reversion by including the pre-1940 level of the outcome variable as a control. Data is obtained from the IPUMS microsample of the 2000 Ghana census. The coefficient on $(Post_{dc} \times Inf_d)$ provides the difference-in-difference estimate comparing the mean outcome variable for younger cohorts (born a decade following the penicillin campaigns) to an older cohort (born a decade prior to the penicillin campaigns) in a high endemicity district, compared to the difference in a low endemicity district. Geographic controls are for latitude, longitude, land area and mean area-weighted slope. Cocoa soil suitability controls is based on an index computed by the Ghana Soil Research Institute, which provides a measure for the overall suitability for cocoa farming. Additional controls are for the distance from the centroid of each district to the nearest railway line, to Accra, to the nearest mission station/hospital, and to the nearest agricultural station. Robust standard errors are in parentheses, and clustered using district administrative boundaries (as at 2000). *** p<0.01, ** p<0.05, * p<0.1.

TABLE 3.6: PLACEBO CHECKS

	(1) Primary school completion	(2) Years of schooling
Panel A: Actual Penicillin Campaign		
$Post_{dc} \times Inf_d$	0.0145***	0.128*
	(0.00525)	(0.0661)
Observations	212,302	212,302
R-squared	0.200	0.189
Panel B: Placebo Penicillin Campaign (Decade Earlier)		
$Post_{dc} \times Inf_d$	0.00606	0.0692
	(0.00871)	(0.107)
Observations	136,490	136,490
R-squared	0.201	0.186
Panel C: Placebo Penicillin Campaign (Decade Later)		
$Post_{dc} \times Inf_d$	0.00766	0.0926
	(0.00674)	(0.0739)
Observations	301,776	301,776
R-squared	0.187	0.181
District fixed effects	YES	YES
Year of birth fixed effects	YES	YES
Geographical controls	YES	YES
Controls for cocoa soil suitability	YES	YES
Controls for dist. to Accra	YES	YES
Controls for dist. to missions/hospitals	YES	YES
Controls for dist. to agricultural station	YES	YES
Controls for mean reversion	YES	YES

NOTES: Panels (A), (B) and (C) respectively report difference-in-difference results based on equation (3.1) for the actual penicillin campaign (1955-1965), for an early placebo campaign (1945-1955) and for a later placebo campaign (1965-1975). Results use data for the female subsample of the IPUMS Ghana 2000 census, and include controls for geographcial characteristics, cocoa soil suitability, proximity to mission stations, hopsitals, agricultural station and to Accra, and controls for mean reversion. Robust standard errors are in parentheses, and clustered using district administrative boundaries (as at 2000). *** $p<0.01$, ** $p<0.05$, *$p<0.1$.

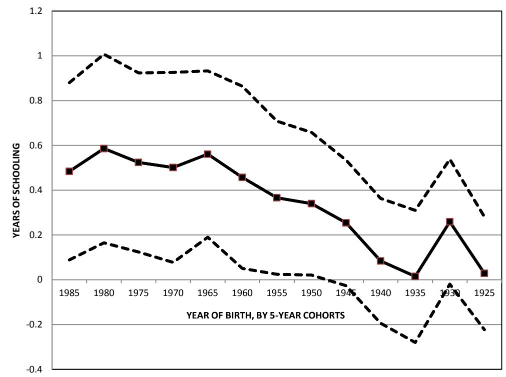

Figure 3.2: Female Years of Schooling. This chart plots the β_c coefficient of $(Post_{dc} \times Inf_d)$ in equation 3.2 for female cohorts born between 1925 and 1985, and grouped into 5-year age bins. Omitted category is for cohorts born prior to 1920. Outcome variable is total years of schooling.

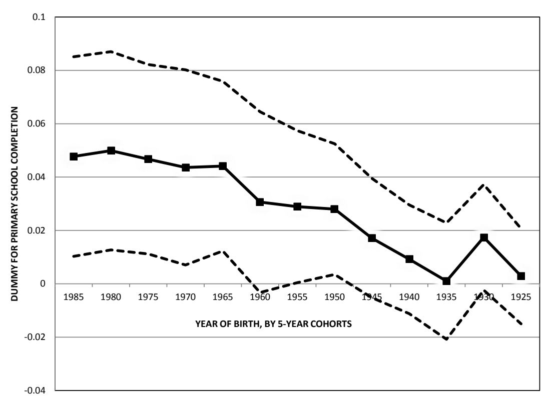

Figure 3.3: Female Primary School Completion. This chart plots the β_c coefficient of $(Post_{dc} \times Inf_d)$ in equation 3.2 for female cohorts born between 1925 and 1985, and grouped into 5-year age bins. Omitted category is for cohorts born prior to 1920. Outcome variable is probability of primary school completion.

3.8 Conclusions

This chapter has examined the impact of the mass treatment campaign of yaws using penicillin which occurred in Ghana over the period 1959-64. The identification exploited the cross-region differences in yaws incidence prior to the campaign, and the fact that the penicillin treatment successfully reduced yaws prevalence in Ghana in the early 1960s. In the main results of this study, I find that a 1 percent decrease in the prevalence of yaws following the penicillin treatment increased the probability of primary school completion, of being literate, and of being able to speak English by 2.1, 1.9 and 2.1 percentage points respectively. In further robustness checks, I observe that the results reported are robust to the inclusion of various time-invariant geographic controls, controls for pre-eradication district characteristics (such as educational attainment, health and agricultural infrastructure) and controls for contemporaneous changes in education and health infrastructure. The results reported in this Chapter motivate the need to re-prioritize eradication of "neglected tropical diseases", given the relatively low cost of such treatment interventions.

Chapter 4

Pests and Politics: Impacts of the Cocoa Swollen Shoot Disease in Ghana

4.1 Introduction

A large literature in empirical social science examines the subject of political partisanship and investigates its formation, its persistence and its implications for political behavior (Campbell et al, 1960; Niemi and Jennings, 1991; Gerber, Huber and Washington, 2010). The traditional debate since the 1970s concerned whether partisanship developed due to affective ties to a social grouping, or rather was the result of instrumental actions influenced by contemporary political events. Motivated by work in the social psychology literature, some scholars viewed party affiliation as the result of political socialization which may be the result of early childhood experiences (for example, within the family or one's immediate community) or due to inherited genetic traits (Campbell et al, 1960; Converse, 1969; Settle et al, 2009). Partisan affiliation may also be the result of a Bayesian updating mechanism based on observations derived from personal experience of political events and politicians' behavior (Fiorina, 1981; Popkin, 1991; Achen, 1992). For example, in a recent contribution, Madestam and Yanagizawa-Drott (2011) investigate the impacts of childhood participation in Fourth of July celebrations in the United States on adult political views. They argue that a higher number of rain-free Fourth of July celebrations in an individual's county during childhood increases the likelihood of voting for the Republican party in adulthood, and results in higher turnout in presidential elections.

Regardless of its mode of formation or transmission, the literature generally finds that partisan affiliations tend to be persistent — over an individual's adult life span, and also across generations (Green, Palmquist, and Schickler, 2002; Niemi and Jennings, 1991). Persistence of partisan loyalties may be partly explained if these affiliations are viewed as inherited norms which serve as heuristics or ”rules-of-thumb” in assessing the platforms of political parties or the quality of individual politicians. The study of partisan affiliation is also of direct importance as it has implications for observed political opinions and behavior (Gerber et al, 2009), as well as for distributive politics (Cox and McCubbins, 1986; Dixit and Londregan, 1995, 1996; Lindbeck and Weibull, 1987).

In this chapter, I contribute to this broad literature by studying the development of partisanship in colonial Ghana, and its implications for distributive politics today. Specifically, I trace the development of political opposition to center-left (Nkrumahist ideology) in southern Ghana to political disturbances which occurred in the late colonial period following outbreaks of the swollen shoot disease which infected cocoa farms. The British colony of the Gold Coast (now Ghana) served as a leading producer of cocoa in the early twentieth century. Following World War II, the cocoa industry in the colony was severely threatened by the outbreak of the swollen shoot disease caused by the cocoa swollen shoot virus (CSSV) and transmitted by mealybugs (*Pseudococccidae*). The disease transmission mechanism has been extensively studied in the cocoa entomology literature: the mealybug vectors are wingless insects, and spread of the disease occurs by radial growth of infected areas and by wind dispersion of the mealybug. The only control means available was to cut down diseased trees wherever outbreaks occurred, resulting in widespread riots in the cocoa-growing regions in the 1940s. The Watson Commission, which was appointed by the colonial government to investigate riots in Ghana sparked by the cocoa swollen shoot disease, confirmed the political motivations of the disturbances (Colonial Office, 1948, pp 49). However, the short- and long-term implications of these disturbances have not been empirically examined.

Using novel historical data from the late colonial period on cultivated cocoa acreages and the spatial variation in CSSV pest infection, I construct a treatment variable of the intensity of the disease shock. The CSSV pest spread in a plausibly random way, and

I provide evidence that the disease treatment measure is uncorrelated with various pre-epidemic district characteristics. Today, based on data from the Afrobarometer surveys, I find that districts which experienced a high intensity of the disease pest report stronger anti-government opinions in survey responses. Specifically, they are more likely to attribute success in life to individual effort rather than government support.

Next, I trace the historical roots of these political views. OLS regression results show that higher pest infection was correlated with higher levels of anti-Nkrumah (center-left) votes in the 1956 Legislative Election — a pivotal election in pre-independence Ghana. These results remain robust to controlling for various pre-epidemic district characteristics such as geographical characteristics, cocoa soil suitability, and proximity to agricultural stations, missionary centers, local hospitals, and transport infrastructure.

Following decades of military rule, the Constitution of the Fourth Republic in Ghana re-introduced multiparty elections and strengthened local government administration by introducing central government transfers to all districts under the *District Assembly Common Fund*. I show that districts which historically experienced high pest infection still vote against the center-left party in both the 1992 and 1996 presidential elections. Using both OLS and 2SLS regressions, I examine the impact of government opposition on district transfers, and find that opposition areas receive lower transfers from the central government during this period. I examine possible violations to the exclusion restrictions of the 2SLS strategy by investigating the impact of the CSSV pest on other economic and social outcomes. I do not find any persistent impacts on cocoa output, educational attainment or economic activity. Using the approach developed by Conley, Hansen and Rossi (2012), I also document that the 2SLS results remain robust to moderate forms of violations to the exclusion restriction assumptions.

This study contributes to two strands of work in the literature. First, as noted earlier, it examines the subject of the formation and persistence of political partisanship. It contributes to the broader literature which examines the impact of personal experience on preferences and behavior. For example, Giuliano and Spilimbergo (2009) argue that individuals who are exposed to macroeconomic recessions in early adulthood are more likely to state preferences in support of state redistribution and view luck rather than effort as important determinants of success. Alesina and Fuchs-Schündeln (2007) also document that

East Germans who lived under a centrally planned economy, have preferences in favor of greater state redistribution. More broadly, the subject of persistence of cultural norms has been previously discussed in the literature (see for example, Salaman (1980), Greif (1994), and Cohen et al., (1996)), and reviewed by Nunn (2009, 2012). Nunn (2012) argues that historical shocks may have persistent effects via a cultural channel, and discusses recent empirical research on cultural persistence. As an example, using data from the Afrobarometer surveys, Nunn and Wantchekon (2011) provide evidence that ethnic groups in Africa which historically experienced high levels of the slave trade report lower levels of trust in others today. Guiso et al. (2008) also observe the persistence of social capital of Italian city states dating from 1000-1300 AD. Tabellini (2008) similarly examines the impact of trust in others and belief in individual effort on comparative economic development in Europe today, using nineteenth century literacy rates and historical political institutions as instruments.

Second, this study also contributes to the literature on distributive politics by examining the patterns of distribution of transfers across districts in Ghana today. Since Laswell (1936), scholars examining the fiscal allocation of government resources among groups have been concerned with one central question: *who gets what, when and how?* Theoretical models of distributive politics have two main predictions. One school of thought argues that rational incumbent politcians behave as risk averse agents, and consequently target resources to loyal constituencies in order to maximize their vote shares (Cox and McCubbins, 1986). An alternate model views incumbent politicians as channelling resources to swing voters who could be pivotal in deciding the outcome of elections (Dixit and Londregan, 1995, 1996; Lindbeck and Weibull, 1987). These models have been empirically tested in various settings. For example, in India, Khemani (2007) observes that intergovernmental allocations made via a central political executive are targeted to politically aligned states, while transfers allocated by an independent body reduce such partisan influences. Similarly, using a panel of US local government transfers for the period 1957-1997, Ansolabehere and Snyder (2002) observe that state governments target transfers to counties which provided the strongest electoral support. In contrast, Dahlberg and Johansen (2002) also provide evidence that electorally pivotal regions in Sweden receive the largest concentration of allocated environmental grants. This paper sheds further light on distributive politics in the case of Ghana, building on earlier contributions by Banful (2009).

The remainder of this chapter proceeds as follows. In section 4.2, I provide a brief historical background examining the growth of the cocoa industry in Ghana in the first half of the twentieth century, and the commencement of the swollen shoot campaigns. Section 4.3 presents the empirical strategy, and in section 4.4, I summarize the data sources used in this paper. Section 4.5 presents OLS and 2SLS results, and section 4.6 discusses possible violations of the 2SLS exclusion restriction assumptions. Conclusions are presented in Section 4.7.

4.2 Historical Background

4.2.1 Cocoa in Ghana (1900-1945)

The British colony of the Gold Coast (now Ghana) served as a leading producer of cocoa in the early twentieth century. Cocoa accounted for about three-quarters of the colony's GDP, and provided an important source of dollar-denominated earnings for the British government. The crop was introduced in southern Ghana in the late nineteenth century in Akwapim by a Ghanaian farmer, Tetteh Quarshie, and the Basel Missionaries. The adoption of this cash crop was widespread in the forest regions of southern Ghana where the soils and climate were best suited for cocoa cultivation. Propagation of cocoa was made possible by seedlings distributed from the Aburi Botanical Gardens and various agricultural stations established in the forest zone.

A striking feature of the growth of the cocoa industry in the Gold Coast was that the colony's cocoa output was dominated by smallholder farmers who depended on cocoa incomes to finance imported goods and other food purchases. In some cases, bands of Akwapim farmers migrated westwards to acquire new forest lands for cocoa cultivation (Hill, 1963). Dickson (1969) surveys the expansion of cocoa cultivation in southern Ghana in the first half of the twentieth century, and observes that without adequate knowledge of the appropriate soil and climatic conditions for cocoa cultivation, the spread of new farms was largely via a process of trial and error (pp 304). The rapid adoption of the crop in the colony was also documented by Cardinall in the 1931 Gold Coast Census:

> So easy is cacao to grow and so remunerative that a false idea of commercial values has been forced upon the native peasant. He has yet to learn to think in pence instead of pounds, and once that lesson has been thoroughly learnt, the only limit to his continued prosperity is that of the supply of labour. (Cardinall, 1932, pp 99).

Prior to World War II, Ghana accounted for about a half of global cocoa production, with the purchase of the local crop dominated by international firms, such as Cadbury and Fry. Export agriculture in the country was essentially cocoa cultivation and many other previous cash crops such as coffee, rubber, cotton and gum copal had reduced economic importance (Dickson, 1969).

4.2.2 Swollen Shoot Campaign (1948-55)

After World War II, however, the cocoa industry in the colony was threatened by the outbreak of the swollen shoot disease. The disease was first reported by local farmers in 1936 around Effiduase, and identified to be a viral infection in 1939. The disease transmission mechanism has been extensively studied in the cocoa entomology literature as being caused by the CSSV virus (genus *badnavirus*) and transmitted by mealybugs (*pseudococccidae*). The mealybug hosts are wingless insects, and spread of the disease occurs by wind dispersion of the insect hosts and radial growth of infected areas (Cornwell, 1958; 1960). Infected trees generally died within 24 months and there was no scientific cure for tackling the plant virus. Thus, wherever outbreaks occurred, the only control means available was to cut down diseased trees as well as immediate neighbors of diseased trees even if these were symptomless. The potential destruction of the colony's cocoa industry created considerable panic for both the colonial government as well as foreign commercial interests which had dominated marketing of West African cocoa (Danquah, 2003; Cocoa Conference, 1951).

The disease was concentrated around the Eastern Region, but with scattered infections all through the cocoa-growing regions. The control measures implemented by the colonial government commenced in 1946 with the cutting-out policy in which state agricultural officials forcibly destroyed infected trees in cocoa growing areas. The policy proved to be extremely unpopular and resulted in widespread protests in southern parts of the country. Consequently, over the period 1948-1951, two major Commissions (the Watson Commision and the 1951 Swollen Shoot Commission) were appointed by the colonial government and the Nkrumah-led administration to investigate the concerns of local farmers and propose remedial actions[1]. The Watson Commission noted the political motivations of the riots in 1948, and reported that ”the Government’s scheme for eradication of the disease is scientifically sound but politically inexpedient” (Colonial Office, 1948, pp 49).

Moreover, there were various rumors associated with the government policy of compulsory tree cutting which spread in the cocoa-growing regions, and which the Watson Commission noted as generating distrust of the central government. Following the riots of 1948, several rumors concerning the future of the country’s cocoa industry circulated in the cocoa belt. For example, the Watson Commission reported comments by local farmers that ”Britain intends to the sell the Gold Coast to the United States but wished to ensure the death of the cocoa industry to avoid subsequent competition” (Colonial Office, 1948, pp 49). The 1951 Commission similarly documented rumors that the cutting-out policy was ”to wipe out the cocoa industry from the Gold Coast”, and ”surveying and demarcation of farms has led to the fear that the farmers’ land is to be confiscated”.

By 1951, the Nkrumah administration halted the cutting-out policy, dissolved the Cocoa Rehabilitation Department, and commenced new programs of grassroots campaigns educating farmers on remedial measures needed to tackle the cocoa swollen shoot disease. Under a revised program, *A New Deal for Cocoa*, the administration re-commenced compulsory cutting-out of infected trees but with compensation paid to farmers for trees cut, and replanting grants also provided to encourage rehabilitation of destroyed farms (Austin, 1964).

4.3 Empirical Strategy

This section presents the empirical strategy adopted in this chapter. Using data from the Afrobarometer Survey, I first examine the relationship between the CSSV disease intensity

[1]First was the Colonial Office (1948), *Commission of Enquiry into Disturbances in the Gold Coast led by Aiken Watson in 1948 [Watson Commission]* and second, the Committee of Enquiry into the Existing Organisation and Methods for the Control of Swollen Shoot Disease by the Compulsory Cutting Out of the Infested Cocoa Trees (1951).

and political views today by running a regression of the form:

$$Y_{id} = \beta_1(High)_d + \beta_2(Medium)_d + \delta Z^{'}_{id} + \gamma X^{'}_d + \epsilon_{id} \tag{4.1}$$

where Y_{id} is the outcome variable of interest, and the vectors $Z^{'}_{id}$ and $X^{'}_d$ respectively contain individual and district level controls. $(High)_d$ and $(Medium)_d$ refer to the fraction of district land area classified as being under high and medium pest infection respectively. I examine the outcome variable related to perceptions of the relative role of individual effort and government support. Specifically, Y_{id} is a dummy equal to one if the individual views personal effort as more important than government support in determining success in life.

I proceed to trace the historical origins of these political views by examining the impacts of the disturbances arising from the cocoa swollen shoot riots on the 1956 Legislative Elections. I run a regression of the form:

$$\begin{aligned} OppVoteShare_d = {} & \beta_1(High \times CocoaShare)_d + \beta_2(Medium \times CocoaShare)_d \\ & + \gamma X^{'}_d + \delta_r + \epsilon_d \end{aligned} \tag{4.2}$$

where $OppVoteShare_d$ refers to the opposition vote share in the 1956 Legislative Election. β_1 and β_2 provide coefficients on the treatment measure for high and medium pest infection areas. The vector $X^{'}_d$ contains various district level controls including time-invariant geographic controls, and distance from the centroid of each cocoa district to the nearest agricultural station, railway station and hospital, as at 1948. δ_r is a region fixed effect, allowing for separate intercepts for the three existing regions in southern Ghana in 1956, namely: the Colony, Ashanti and Transvolta Togoland. The district borders employed here refer to electoral boundaries for the 1956 Gold Coast Legislative Assembly Elections (Austin, 1964, pp 204).

Next, the instrumental variable 2SLS strategy of the impact of opposition vote share on government transfers is subsequently implemented as follows. Specifically, I use the CSSV pest intensity interacted with the share of total acreage under cocoa cultivation (in 1947) as an instrument for the opposition vote share, controlling for the district mean area-weighted cocoa suitability index, and other geographic characteristics. I examine the first stage relationship between historial disease intensity and opposition vote share today in the regression:

$$OppVoteShare_d = \beta_1(High \times CocoaShare)_d + \beta_2(Medium \times CocoaShare)_d + \gamma X_d^{'} + \delta_r + \epsilon_d \tag{4.3}$$

where $OppVoteShare_d$ refers to the opposition vote share in the 1992 or 1996 Presidential Elections. Next, I examine the impact of the oppostion vote share on government transfers as in the regression below:

$$GovtTransfer_d = \alpha OppVoteShare_d + \gamma X_d^{'} + \delta_r + \epsilon_d \tag{4.4}$$

where $GovtTransfer_d$ refers to central government transfers from the District Assembly Common Fund to district, d. Here, I use the CSSV intensity measure as an instrument for the opposition vote share.

This instrumental variable (IV) strategy relies on two conditions (Angrist and Pischke, 2008). First, the CSSV treatment measure must be correlated with the opposition vote share, i.e. the existence of a first stage relationship. If the first stage relationship is weak, then the IV estimates are likely to be noisy, and biased towards the OLS results (Bound, Jaeger, and Baker 1995; Staiger and Stock 1997). The second requirement of the IV strategy is the exclusion restriction which requires that the CSSV pest treatment be uncorrelated with other determinants of the outcome measure of interest, i.e. $corr(CSSVTrmt_d, \epsilon_d) = 0$. This requirement is comprised of two parts: that the CSSV treatment be as good as randomly assigned conditional on covariates, and that the treatment measure affects government transfers today only via the opposition vote share channel.

The exclusion restriction is not directly testable. However, I examine its plausibility in a number of ways. The historical accounts indicate that the CSSV disease propagation was plausibly random following wind dispersal of the insect hosts. I confirm the lack of correlation between the CSSV shock and pre-existing district characteristics by regressing various pre-epidemic district characteristics obtained from the 1948 Gold Coast Population Census on the CSSV intensity measure (see Table 4.1B). In section 4.6, I also present further checks for violations of the exclusion restrictions by examining the impacts of the pests on other outcomes such as cocoa output, educational attainment and employment. Finally, I implement the approach proposed by Conley, Hansen and Rossi (2002) to examine the sensitivity of my IV/2SLS results to violations of the assumptions of the exclusion restrictions.

4.4 Data

Data on cocoa swollen shoot virus (CSSV) prevalence is obtained from maps prepared by the Gold Coast Agricultural Department and presented at the conferences of the Cocoa, Chocolate, and Confectionery Alliance (1957), see Figure 4.1. The map provides five levels of increasing intensity of the pest infection: areas with no recorded infections; areas with few small scattered outbreaks; areas with numerous, large, scattered outbreaks; devastated areas; and abandoned areas. Data on cultivated cocoa areas circa 1947 is obtained from the *Report on the Progress of Swollen Shoot Control*, see Figure 4.2. I construct a measure of the disease intensity by superimposing district boundaries (from the 1948 Gold Coast census) on maps of the CSSV prevalence and cultivated cocoa areas. I consider two measures of the treatment. The first measure examines the fraction of a district's landarea which is classified as under high, medium or low CSSV infestation. I aggregate areas with large outbreaks, devastated areas and abandoned areas as *high* treatment intensity; and areas with few scattered outbreaks as *medium* intensity areas; and uninfected areas as *low* treatment areas.

I also consider a second measure of the treatment intensity in which I interact the CSSV disease severity with the fraction of land under cocoa cultivation in 1947. The observations from Cardinall (1931) and Dickson (1969) above suggest that the cultivation of the crop was extensive in many parts of southern Ghana, but without much understanding of the soil and climactic suitability of specific locations. The second treatment measure provides an index of the relative severity of the disease across districts in the cocoa belt by weighting the disease severity with the fraction of land area under cocoa cultivation in 1947. In using this second measure, I also control for the mean area-weighted cocoa suitability index for the district. The cocoa suitability index is computed by the Ghana Soil Research Institute and provides a measure for the overall suitability for cocoa farming on a GIS grid raster. The index provides a 0-100 score of cocoa suitability for each grid cell, taking into account local precipitation, temperature and soil quality (see Figure 4.3).

Data on pre-treatment district population and economic characteristics are obtained from the *Gold Coast Census of 1948*. For post-treatment education and labor force outcomes, I use the IPUMS microsample from the 2000 Census of Ghana. Data on geographic characteristics is obtained by overlaying the 1948 administrative map of Ghana/Gold Coast on elevation and slope data obtained from SRTM/ArcMaps and computing the mean area-weighted elevation and slope. I also calculate various district-level geographic controls, including the distance from the centroid of each district to the nearest railway line and major road, to the nearest protestant mission, and to the nearest hospital in 1948.

I use data from the Afrobarometer Surveys for Ghana to examine questions related to perceptions of the role of individual effort versus government support in attaining personal success. The Afrobarometer Surveys are nationally representative surveys of social and political opinions of voting age citizens in African countries, and used in previous research in the literature (see Nunn and Wantchekon, 2009). I utilize data from Rounds 2 and 3 of the Ghana surveys, where the question on the relative importance of individual effort versus government support was posed to respondents.[2]

Data for outcomes of the 1956 Gold Coast Legislative Elections is obtained from various issues of the *Daily Graphic* newspaper from July 1956. From these sources, I calculate the combined opposition vote shares as $(1 - CPPVoteShare)$[3]. Electoral outcomes data for the 1992 and 1996 presidential elections in Ghana are obtained from the Ghana Elec-

---

[2]Specifically, the outcome variable here is coded from the following Afrobarometer question (see Bauer (2005), Afrobarometer Data Codebook for Ghana (Round 2, Q61), pp 41): *Lets talk for a moment about the kind of society we would like to have in this country. Which of the following statements is closest to your view? Choose Statement A or Statement B. (A). People should look after themselves and be responsible for their own success in life. (B). The government should bear the main responsibility for the well-being of people.*

[3]Nkrumah's Conventions Peoples Party (CPP) won 57 percent of the overall votes cast, but faced opposition from other parties viz.: NLM (National Liberation Movement), MAP (Muslim Association Party), NPP (Northern People's Party), TC (Togoland Congress), FYO (Federated Youth Organization), and the WYA (Wassaw Youth Association) — see Austin (1964) pp 354.

Figure 1: Cocoa Swollen Shoot Disease, 1948

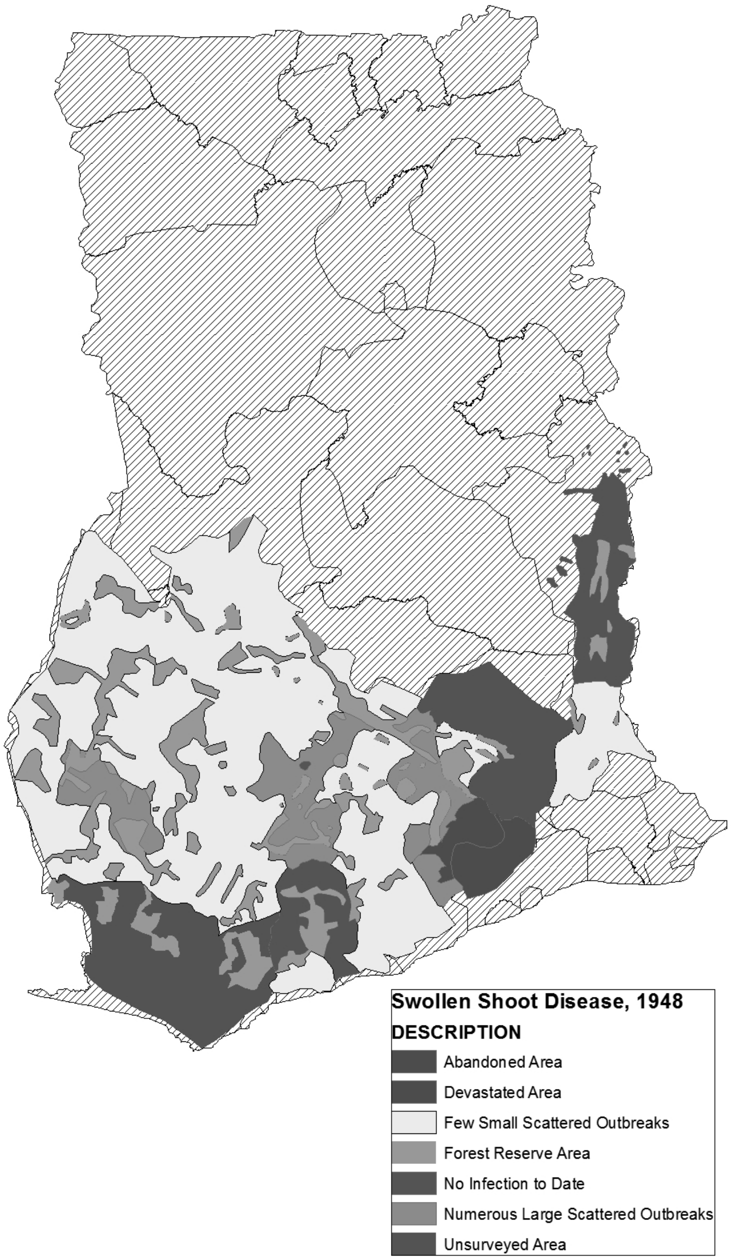

Figure 4.1: Source: Cocoa, Chocolate, and Confectionery Alliance Conference (1957)

Figure 2: Gold Coast/Ghana Cocoa Areas, September 1947

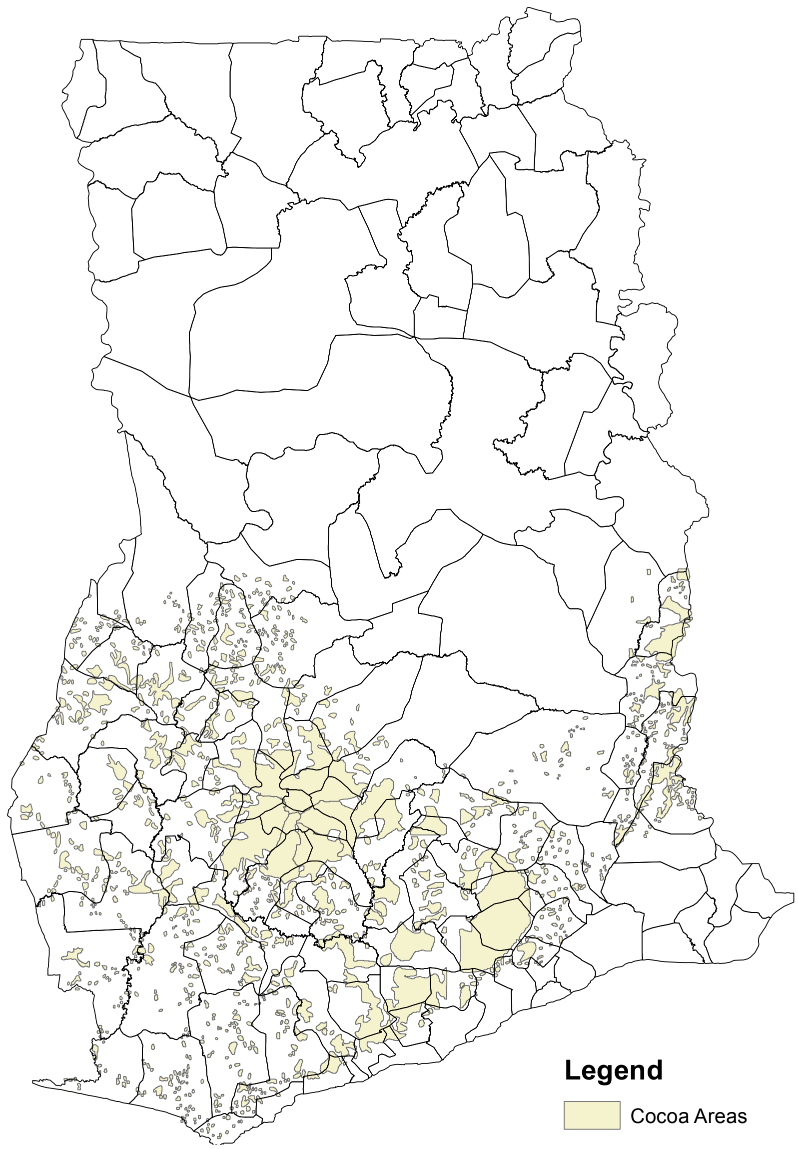

Figure 4.2: Source: Report on the Progress of Swollen Shoot Control

Figure 3: Cocoa Soil Suitability Index

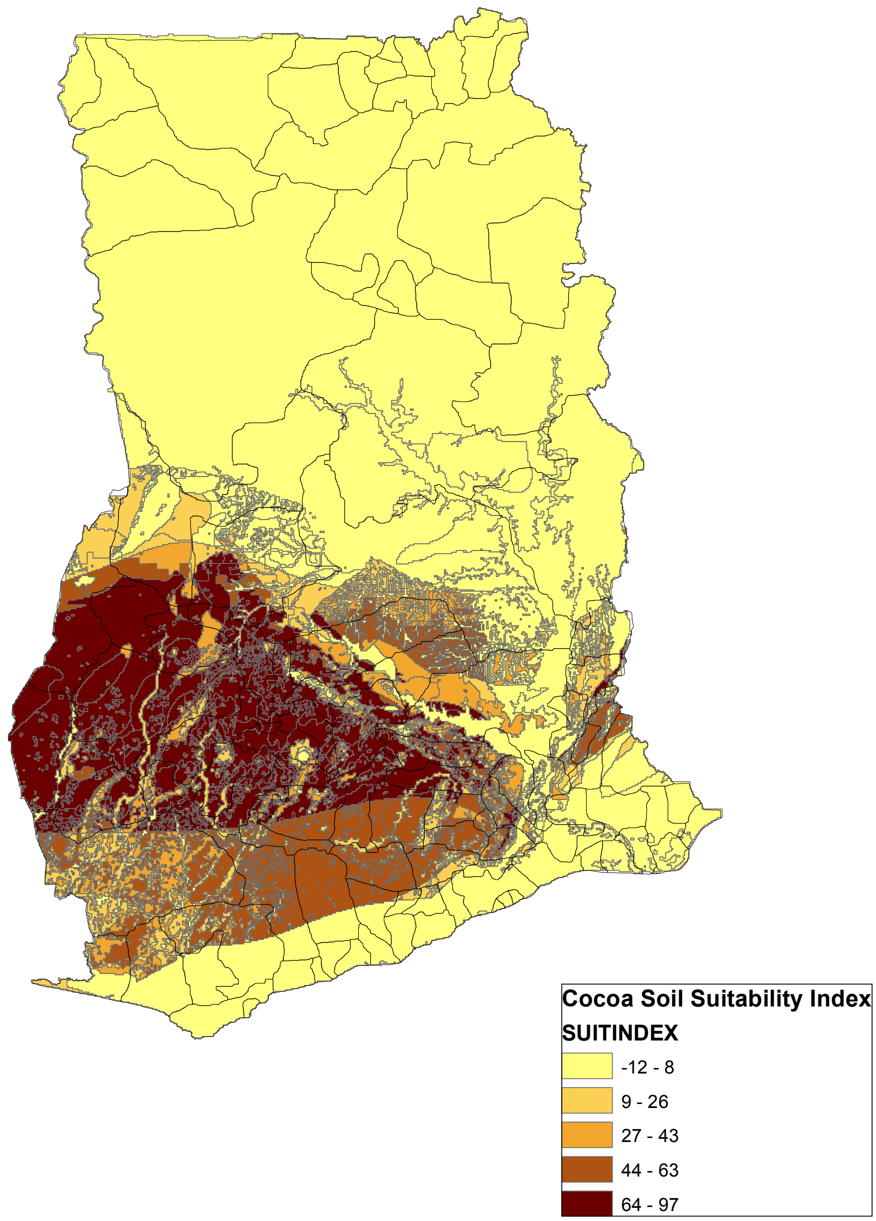

Figure 4.3: Source: Boateng et al., (1999), Soil Research Institute, Ghana

toral Commission, while data on national government transfers to districts for the period 1994-2000 are obtained from the Ghana District Assembly Common Fund (see Banful (2009)). Data limitations imply that I am able only to examine 2 elections in the 1990s.

Finally, I compile data on cocoa output spanning the period 1927-2007 from a variety of sources. Data is compiled for the years 1927, 1936, 1947-1965, 1969-1973, 2003-2007 as follows: from Dickson (1969, pp 168-9) for 1927 and 1936; Analysis of Cocoa Purchases by Societies, Districts and Regions for 1947-1965; the Ghana Cocoa Marketing Board Newsletter for 1969-1973; and the Ghana COCOBOD district-level purchases for the period 2003-2007. The Ghana COCOBOD demarcates the cocoa-growing areas in Ghana into 7 *cocoa regions*: Eastern, Ashanti, Brong-Ahafo, Central, Volta, Western North and Western South, which are subdivided into *cocoa districts* distinct from administrative districts. The data spanning most decades in the twentieth century are organized into 28 cocoa districts.

Table 4.1A provides summary statistics for various outcome variables and covariates used in this paper, comparing cocoa and non-cocoa districts. In Table 4.1B, I regress various pre-epidemic district characteristics on the disease intensity variables. The disease intensity measure is orthogonal to most district characteristics, the only exceptions being population density in column (1). This is potentially a concern for my identification strategy as initially high population density may be correlated with other factors (such as income) which influence my outcome variables. In subsequent regressions, I therefore control directly for pre-epidemic population density. The significant correlation in column (1) is however driven result of two influential observations in the highly devastated districts in the Eastern Region, and the correlation disappears when I omit these two districts. In specifications using the disease treatment interacted with the share of cocoa acreage, I also control for distance to railways and to church missions.

TABLE 4.1A: SUMMARY STATISTICS

	All districts (1)	Cocoa districts (2)
Fraction of land area under high pest infection	...	0.161
		(0.279)
Fraction of land area under medium pest infection	...	0.419
		(0.349)
Fraction of district land area under forest reservation	0.0517	0.0837
	(0.0885)	(0.100)
Population (1948 Census)	59,585	54,181
	(33,605)	(26,383)
Population density (persons per sq. km., 1948 Census)	74.11	76.96
	(106.2)	(94.58)
Fraction of population employed in agriculture (1948 Census)	0.712	0.677
	(0.132)	(0.0748)
Fraction of population employed in mining (1948 Census)	0.0564	0.0810
	(0.159)	(0.186)
Fraction of population employed in manufacturing (1948 Census)	0.0986	0.112
	(0.0477)	(0.0355)
Local government revenue (Sh. per capita, 1948 Census)	0.476	0.522
	(0.501)	(0.524)
Cultivated cocoa acreage per capita (sq. km per capita)	...	0.0081
		(0.0068)
Land area (sq. km.)	3,449	2,812
	(4,500)	(3,456)
Latitude	7.190	6.267
	(1.912)	(0.781)
Longitude	-0.858	-1.053
	(0.971)	(0.909)
Mean area-weighted elevation (in meters)	155.8	166.8
	(90.31)	(85.28)
Mean area-weighted slope (in degrees)	2.639	3.492
	(1.449)	(1.185)
Cocoa soil suitability index	20.35	33.19
	(27.51)	(28.20)
Distance to railway (in km)	138.2	47.6
	(161.8)	(45.0)
Distance to protestant mission (in km)	72.6	41.9
	(72.8)	(35.4)
Distance to hospital (in km)	18.8	19.3
	(11.1)	(11.2)
CPP vote share (1956 Legislative Elections)	62.10	66.10
	(25.10)	(25.30)
Government vote share (1992 Presidential Elections)	58.77	57.46
	(18.50)	(18.31)
Government vote share (1996 Presidential Elections)	61.56	56.98
	(19.57)	(19.01)
Log per capita government transfers (1996)	8.011	8.035
	(0.343)	(0.321)
Log per capita government transfers (2000)	9.298	9.288
	(0.417)	(0.394)
Observations	102	63

NOTES: Cocoa districts defined as districts with cocoa farms as at September, 1947. Data on intensity of cocoa swollen shoot outcomes are obtained from reports of the *Cocoa, Chocolate, and Confectionery Alliance Conference (1957)*, see Fig. 1. Population data and employment shares in various sectors are obtained from the 1948 Gold Coast Census. Data on geographic characteristics obtained by overlaying administrative map of Ghana/Gold Coast on elevation (slope) data obtained from SRTM, and computing mean area-weighted elevation (slope). Data on CPP vote shares in 1956 Legislative Elections is obtained from *Daily Graphic* newspapers of July 1956. Data on government vote shares and transfers to local governments (1992, 1996, and 2000) is obtained from Banful (2009).

TABLE 4.1B: PRE-EPIDEMIC DISTRICT CHARACTERISTICS (1948)

VARIABLES	Population Density (1)	Cocoa acreage per capita (2)	Frac. pop. in agric. (3)	Frac. pop. in mining (4)	Frac. pop. in manuf. (5)	Dist. to railways (6)	Dist. to church mission (7)	Dist. to hospital (8)
FractionHighPest	87.28*	0.00122	-0.0481	-0.0386	0.00405	-25,326	3,839	-9,623
	(45.68)	(0.00350)	(0.0452)	(0.104)	(0.0197)	(16,663)	(15,645)	(6,307)
FractionMediumPest	74.10	0.00242	0.0133	-0.0130	-0.00803	931.9	-9,208	-7,397
	(47.20)	(0.00278)	(0.0405)	(0.0840)	(0.0151)	(10,192)	(9,064)	(4,818)
FractionHighPest * ShareCocoaArea	132.8	0.00422	-0.0989	-0.0894	0.0123	-44,657	-7,854	-10,104
	(109.2)	(0.00661)	(0.0614)	(0.0903)	(0.0376)	(39,248)	(19,733)	(8,856)
FractionMediumPest * ShareCocoaArea	171.0*	0.000224	-0.00160	0.0533	-0.00132	-31,312**	-27,188*	-5,205
	(100.3)	(0.00517)	(0.0369)	(0.0927)	(0.0205)	(13,598)	(15,517)	(6,320)
Observations	48	48	48	48	48	48	48	48

NOTES: All columns include geographical controls for latitude, longitude, slope and altitude. The dependent variable in column (1) is district population density (1948 census). In column 2 the dependent variable is the cocoa acreage per capita as at 1947 (see Figure 2). The dependent variables in columns 3, 4, and 5 refer respectively to the fraction of the population employed in agriculture, in mining, and in manufacturing from the 1948 census. In column 6, the dependent variable is the local government revenue per capita in 1948. For columns 7, 8 and 9, the dependent variable refers to the distance from the centroid of each district to the nearest railway line, to the nearest protestant mission, and to the nearest hospital. Robust standard errors are in parentheses. *** p<0.01, ** p<0.05, * p<0.1

4.5 Results

4.5.1 OLS Results

In this subsection, I report results on the impacts of the disturbances following the CSSV on political outcomes using data from the Afrobarometer survey and the 1956 Gold Coast Legislative Elections. Table 4.2 presents results of the relationship between the CSSV pests intensity and the dummy for whether individual effort rather than government support is important for personal success. In column (1), I present the baseline correlation between the outcome variable and the CSSV treatment intensity, controlling only for individual characteristics (age and gender). In column (2), I include time invariant geographic controls (for latitude, slope and cocoa soil suitability). Columns (3) to (5) successively include controls for distance from the centroid of each district to the nearest cocoa stations, agricultural stations and railway line as at 1948. Considering columns (3) to (5), there is a robust, positive and significant relationship between the CSSV intensity variable and support for the view about the importance of individual effort (rather than government support) in determining personal success.

Table 4.3 documents the impact of the CSSV treatment intensity on the outcomes of the Gold Coast 1956 Legislative Elections controlling for various pre-treatment characteristics. Robust standard errors are presented in brackets. Given the spatial clustering of the CSSV pests, I also report Conley standard errors in square brackets, which correct for two-dimensional spatial dependence. The Conley covariance matrix provides a weighted average of spatial autocovariances, with the weights starting at one and decaying linearly to zero within a cutoff region. I set the cutoff to 2 degrees (or roughly 220 kilometers) in both longitude and latitude dimensions. Column (1) examines the baseline relationship between the CSSV disease intensity and opposition vote share in 1956. Columns (2) to (6) present results with region fixed effects, geographic controls, and additional controls for proximity to the nearest agricultural station, railway station and hospital infrastructure in 1948. There is a robust and statistically significant relationship for specifications in columns (2) to (6). Using the preferred specification in column (3), a one standard deviation increase in the high pest treatment measure results in a 4 percentage point increase in the opposition vote share. Similarly, a one standard deviation increase in the medium pest treatment measure increases opposition vote share by 5 percentage points.

4.5.2 2SLS Results

In this subsection, I use an instrumental variables approach to examine the impact of supporting the government opposition on transfers received from the District Assembly Common Fund (DACF) in Ghana. Following the 1992 Ghanaian Constitution, the DACF was established to decentralize public expenditures and strengthen local government administration. Transfers from the DACF forms nearly 80 percent of the total expenditures of local governments. Although DACF allocations to local governments are formula-based, the disbursement formula has been subject to revisions annually since 1994. These changes arise largely from changes in the allocation criteria and relative weights used in computing the size of transfers to various districts. Banful (2009) observes that allocations from the DACF in Ghana over the period 1994-2000 were subject to political manipulation, with transfers being targeted to reward loyal constituencies. This follows one strand of the theoretical predictions in the distributive politics literature in which elected political parties behave as risk-averse investors and target resources at loyal supporters (Cox and McCubbins, 1986).

Table 4.4 reports the relationship between the CSSV intensity measure and electoral outcomes from 1992 and 1996 Presidential Elections. This serves as the first stage of the 2SLS strategy. Column (1) presents the baseline unconditional relationship between the opposition vote shares and the various treatment measures. Columns (2) to (6) provide additional specifications with region fixed effects, and controlling for geographical characteristics, economic activity, proximity to agricultural stations, mission stations, and hospital infrastructure, and for local government revenues in 1948. Again, we observe a robust and statistically significant relationship between the opposition vote share today and the treatment measure of CSSV pest intensity. For the 1992 presidential elections, the results from the preferred specification in column (3) indicate that a one standard deviation increase in the high pest intensity treatment measure increases the opposition vote share today by about 3 percentage points, while the corresponding increase in medium pest treatment variable increased the opposition vote share by 6 percentage points.

There is typically a first stage relationship between the measure of the CSSV pest treatment and opposition vote share today with the F-statistic ranging from 6.71 to 13.98 for the 1996 election results. In the preferred specification in column (3), the F-statistic equals 13.98, falling within the range needed to avoid weak instrument concerns (Stock, Wright, and Yogo; 2002).

Table 4.5 reports the OLS and IV estimates of the effect of opposition vote shares on the magnitude of DACF allocations. The dependent variable is the mean log per capita government allocations for the period 1997-99 obtained from the District Assembly Common Fund. All specifications include geographic controls and region fixed effects which provide separate intercepts for the 3 colonial administrative regions in the cocoa-growing regions in 1956. Specifications in columns (2) to (6), successively introduce controls for economic activity (share of the population in manufacturing and in mining form the 1948 Census), controls for proximity to agricultural stations, to mission stations and to hospital infrastructure, and for local government revenues as at 1948.

The OLS results presented in Table 4.5, Panel A provides a negative relationship between support for the opposition parties and DACF transfers. Consistent with our hypotheses, the estimates for *oppvoteshare* are all negative, and statistically significant in specifications (1), (2), (3) and (6). The point estimates range from -0.59 to -0.72, implying that a one standard deviation increase in opposition vote share reduces district per capita DACF allocations between 11.2 and 13.6 percent. Next, Table 4.5, Panel B reports the results of the relationship between opposition vote share and DACF transfers using the IV strategy. This is the second-stage results, with the corresponding first stage results previously reported in Table 4.4. The IV point estimates are larger in magnitude than the corresponding OLS results, and range between -1.19 and -1.49. Thus, a one standard deviation increase in opposition vote share reduces per capita government transfers between 22.4 and 28.2 percent. The increase in the magnitude of the point estimate may be partly explained by an upward bias in the OLS results due to omitted variables.

TABLE 4.2: INDIVIDUAL EFFORT VERSUS GOVERNMENT SUPPORT

	Dependent variable: *Dummy for whether respondent believes individual effort is more important than government support for personal success (mean = 0.566)*				
	(1)	(2)	(3)	(4)	(5)
FractionHighPest	0.0739	0.0729	0.140**	0.146**	0.150**
	(0.0541)	(0.0587)	(0.0678)	(0.0683)	(0.0718)
FractionMediumPest	0.179***	0.194***	0.242***	0.242***	0.245***
	(0.0428)	(0.0368)	(0.0518)	(0.0511)	(0.0513)
Observations	1,431	1,431	1,431	1,431	1,431
R-squared	0.013	0.020	0.021	0.021	0.021
Geographical controls	NO	YES	YES	YES	YES
Controls for dist. to cocoa stations	NO	NO	YES	YES	YES
Controls for dist. to agricultural stations	NO	NO	NO	YES	YES
Controls for dist. to railway	NO	NO	NO	NO	YES

NOTES: Dependent variable is a dummy for whether respondent believes individual effort rather than government support is important for personal success (see Bauer (2005), Afrobarometer Data Codebook for Ghana (Round 2, Q61), pp 41). Controls are for district geographical characteristics, and proximity to cocoa stations, agricultural stations and nearest railway line. *** $p<0.01$, ** $p<0.05$, * $p<0.1$

TABLE 4.3: GOLD COAST LEGISLATIVE ELECTIONS (1956)

	(1)	(2)	(3)	(4)	(5)	(6)
Dependent variable: *OppositionVoteShare*						
FractionHighPest * ShareCocoaArea	0.200	0.296**	0.276**	0.319**	0.261*	0.255*
	(0.162)	(0.127)	(0.135)	(0.145)	(0.133)	(0.136)
	[0.147]	[0.050]	[0.059]	[0.065]	[0.070]	[0.071]
FractionMediumPest * ShareCocoaArea	0.536***	0.204**	0.244**	0.298**	0.234*	0.229*
	(0.143)	(0.0951)	(0.110)	(0.127)	(0.124)	(0.124)
	[0.093]	[0.055]	[0.063]	[0.070]	[0.068	[0.065]
Geographical Controls	NO	NO	YES	YES	YES	YES
Controls for Agricultural stations (1948)	NO	NO	NO	YES	YES	YES
Controls for Railway infrastructure	NO	NO	NO	NO	YES	YES
Controls for Hospital infrastructure	NO	NO	NO	NO	NO	YES
Region Fixed Effects	NO	YES	YES	YES	YES	YES
Observations	66	66	66	66	66	66
R-squared	0.190	0.580	0.589	0.596	0.620	0.620

NOTES: Dependent variable is the opposition vote share, i.e. (1 - CPP voteshare) in the 1956 Gold Coast/Ghana Legislative Elections. *FractionHighPest* and *FractionMediumPest* respectively refer to the fraction of a district's land area classified as exposed to high and medium pest infection respectively. *ShareCocoaArea* is the fraction of district's land area under cocoa cultivation as at 1947. District boundaries refer to legislative assembly electoral boundaries (see Austin 1964, pp 204). Region fixed effects provide separate intercepts for the 3 administrative regions in the cocoa-growing regions in 1956, namely: the Colony, Ashanti, and Transvolta Togoland. Geographical controls are for latitude, slope and cocoa soil suitability characteristics. Controls for agricultural stations, railway infrastructure and hospital infrastructure are respectively controls for the distance from centroid of district to nearest agricultural station, railway line and hospital as at 1948. Robust standard errors are in parentheses. Conley standard errors, correcting for spatial dependence, are in square brackets. *** $p<0.01$, ** $p<0.05$, * $p<0.1$

4.6 Robustness Checks

The exclusion restrictions for the IV strategy require the CSSV pest treatment to be uncorrelated with other determinants of the outcome measure of interest. In particular, we expect the treatment measure to affect government transfers today only via the opposition vote share channel. However, the requirement of perfect exogeneity of the instrument is a strict condition, which in many instances, is unlikely to hold exactly. For example, it is plausible that the CSSV pest shock of the 1940s altered the agricultural endowment of the cocoa districts and thus led to changes in cocoa ouput and prevailing economic activity in these districts. The historical accounts however suggest that the impact of the CSSV pest was temporary, and the major cocoa-related Commissions of the period provided various re-planting grants to farmers in affected areas (see section 4.2.2).

In this section, I provide two checks to examine plausibility of the exclusion restrictions. First, using a panel dataset of cocoa output spanning the period 1927-2007, I investigate the response of cocoa output to the CSSV pest. In addition, I retrospectively examine the evolution of education outcomes and economic activity following the CSSV pest using the IPUMS microsample of the 2000 Ghana Census. Second, I adopt an approach developed by Conley, Hansen and Rossi (2012) in which inference is conducted but with gradual relaxation of the exclusion restrictions. This approach enables us to obtain bounds on the effect of the CSSV pest instrument even when we deviate from the assumption of perfect exogeneity.

4.6.1 Cocoa Output

In this section, I examine the evolution of cocoa output since the early twentieth century using panel data covering all 28 cocoa districts in southern Ghana[4]. I examine the impact of the CSSV pest on successive cross-sections of cocoa output data for the 80-year period

[4] As noted earlier in section 4.4, boundaries of cocoa districts differ from regular administrative districts.

TABLE 4.4: PRESIDENTIAL ELECTIONS (1992, 1996)

	(1)	(2)	(3)	(4)	(5)	(6)
			Dependent variable: *OppositionVoteShare1992*			
FractionHighPest * ShareCocoaArea	0.204*	0.208*	0.215*	0.222*	0.226*	0.191
	(0.118)	(0.121)	(0.126)	(0.129)	(0.133)	(0.133)
	[0.063]	[0.062]	[0.066]	[0.071]	[0.072]	[0.072]
FractionMediumPest * ShareCocoaArea	0.302***	0.298***	0.296***	0.230***	0.230***	0.227***
	(0.0668)	(0.0681)	(0.0690)	(0.0724)	(0.0736)	(0.0742)
	[0.055]	[0.057]	[0.057]	[0.051]	[0.051]	[0.051]
			Dependent variable: *OppositionVoteShare1996*			
FractionHighPest * ShareCocoaArea	0.200*	0.204*	0.223*	0.232*	0.225*	0.197
	(0.103)	(0.112)	(0.120)	(0.125)	(0.129)	(0.130)
	[0.059]	[0.061]	[0.066]	[0.074]	[0.077]	[0.076]
FractionMediumPest * ShareCocoaArea	0.298***	0.307***	0.302***	0.220***	0.221***	0.218***
	(0.0611)	(0.0634)	(0.0630)	(0.0653)	(0.0656)	(0.0663)
	[0.045]	[0.046]	[0.045]	[0.035]	[0.035]	[0.034]
F-stat	13.27	13.35	13.98	7.89	7.84	7.25
Geographical Controls	YES	YES	YES	YES	YES	YES
Economic activity (1948)	NO	YES	YES	YES	YES	YES
Controls for agricultural stations	NO	NO	YES	YES	YES	YES
Controls for mission stations	NO	NO	NO	YES	YES	YES
Controls for hospital infrastructure	NO	NO	NO	NO	YES	YES
Controls for local government revenues (1948)	NO	NO	NO	NO	NO	YES
Region fixed effects	YES	YES	YES	YES	YES	YES
Observations	77	77	77	77	77	77
R-squared	0.344	0.580	0.643	0.643	0.661	0.663

NOTES: Dependent variable is the opposition vote share in 1992 or 1996 Presidential Elections. *FractionHighPest* and *FractionMediumPest* respectively refer to the fraction of a district's land area classified as exposed to high and medium pest infection respectively. *ShareCocoaArea* is the fraction of district's land area under cocoa cultivation as at 1947. Region fixed effects provide separate intercepts for the 3 colonial administrative regions in the cocoa-growing regions in 1956, namely: the Colony, Ashanti, and Transvolta Togoland. Geographical controls are for longitude, latitude, slope and landarea. Controls for economic activity are for the share of the population in manufacturing and in mining (1948 Census). Controls for agricultural stations controls for distance to nearest agricultural station. Controls for mission station and hospital infrastructure controls for distance to nearest protestant mission and hospital as at 1948. Robust standard errors are in parentheses. Conley standard errors, correcting for spatial dependence, are in square brackets. *** p<0.01, ** p<0.05, * p<0.1

TABLE 4.5: 2SLS RESULTS

	(1)	(2)	(3)	(4)	(5)	(6)
Dependent variable: *Log per capita government allocations (1997-1999)*						
PANEL A: OLS RESULTS						
OppVoteShare1996	-0.669*	-0.678*	-0.721**	-0.646	-0.597	-0.694*
	(0.355)	(0.353)	(0.358)	(0.399)	(0.391)	(0.402)
Observations	77	77	77	77	77	77
R-squared	0.143	0.148	0.172	0.175	0.203	0.226
PANEL B: 2SLS RESULTS						
OppVoteShare1996	-1.192*	-1.214**	-1.196**	-1.344*	-1.212*	-1.490*
	(0.627)	(0.604)	(0.578)	(0.724)	(0.691)	(0.857)
Observations	77	77	77	77	77	77
R-squared	0.109	0.114	0.145	0.128	0.167	0.168
Geographical Controls	YES	YES	YES	YES	YES	YES
Economic activity (1948)	NO	YES	YES	YES	YES	YES
Controls for agricultural stations	NO	NO	YES	YES	YES	YES
Controls for mission stations	NO	NO	NO	YES	YES	YES
Controls for hospital infrastructure	NO	NO	NO	NO	YES	YES
Controls for local government revenues (1948)	NO	NO	NO	NO	NO	YES
Region fixed effects	YES	YES	YES	YES	YES	YES

NOTES: Dependent variable is the mean log per capita government allocations for the period 1997-99, obtained from the District Assembly Common Fund. Region fixed effects provide separate intercepts for the 3 colonial administrative regions in the cocoa-growing regions in 1956, namely: the Colony, Ashanti, and Transvolta Togoland. Geographical controls are for longitude, latitude, slope and landarea. Controls for economic activity are for the share of the population in manufacturing and in mining (in 1948). Controls for agricultural stations controls for distance to nearest agricultural station. Controls for mission station and hospital infrastructure controls for distance to nearest protestant mission and hospital as at 1948. Robust standard errors are in parentheses. *** $p<0.01$, ** $p<0.05$, * $p<0.1$

spanning 1927 to 2007 by running the regression:

$$
\begin{aligned}
C_{dt} = & \sum_{t=1949}^{2007} [(High \times CocoaShare)_d \times \phi_t)]\beta_{1t} \\
& + \sum_{t=1949}^{2007} [(Medium \times CocoaShare)_d \times \phi_t)]\beta_{2t} \\
& + \gamma(X_d' \times \phi_t) + \delta_d + \delta_t + \epsilon_{dt}
\end{aligned}
\tag{4.5}
$$

where C_{dt} refers to cocoa output in district, d, in year, t; X_d' is a vector containing district level controls; δ_d and δ_t refer to district and year fixed effects and ϕ_t is a dummy for year t. As before, the CSSV treatment variable is given by the fraction of district land area classified as high or medium pest intensity interacted with the fraction of cultivated cocoa area as at *1947*. Here, I am interested in how the coefficients β_{1t} and β_{2t} vary over time. The regression above provides estimates for the beta coefficients for 15-year observations between 1949 and 2007. The omitted categories are cocoa output for four observations which preceded the cutting-out policy: for 1927, 1936, 1947 and 1948.

Table 4.A1 in the appendix provides the coefficients β_{1t} and β_{2t}. There is some evidence that the impact of the shock was temporary: cocoa output initially declines in areas which experienced a high intensity of the CSSV pest. The magnitude of β_1 coefficient is negative and significant up to the late 1950s. Subsquently, the point estimate remains negative although the 90 percent confidence intervals now include zero. For areas which experienced a medium intensity of the pest, the β_2 coefficient is never significant, and in some instances the point estimate is positive. The negative magnitude for the β_1 coefficient appears to be driven largely by one influential observation in the ”abandoned” region around Koforidua. Excluding the abandoned region shows even weaker effects of the CSSV pest on cocoa output. Overall, this account is consistent with qualitative evidence which document that following the initial political disturbances, the CSSV disease was largely contained. With the introduction of new disease-resistant cocoa varieties, and financial support via replanting grants paid to local farmers, the effect of the pest infestation on cocoa output was only temporary.

4.6.2 Education and Employment Outcomes

The exclusion restrictions would also be violated if the impact of the pests in the 1950s resulted in changes in educational attainment in cocoa districts, or the reallocation of labor from agriculture to the manufacturing or service sectors. Detailed panel data on educational attainment and employment outcomes (for example, from waves of the Ghanaian census), are not available. However, I am able to shed light on changes in educational attainment and employment activity by retrospectively examining microsample data of the 2000 Ghanaian Census. The IPUMS microsample for the Ghanaian Census provides data on educational attainment and employment activity for about 1.8 million individuals. Under the assumption that the district of current residence (observed in the 2000 Census) is highly correlated with the district of birth and education, I am able to examine changes in educational attainment and employment activity across various year-of-birth cohorts in districts exposed to varying intensity of the CSSV disease pest.

As in equation (4.5) above, I run the empirical specification below:

$$\begin{aligned} Y_{dt} = &\sum_{t=1}^{10}[(High \times CocoaShare)_d \times \phi_t)]\beta_{1t} \\ &+ \sum_{t=1}^{10}[(Medium \times CocoaShare)_d \times \phi_t)]\beta_{2t} \\ &+ \gamma(X_d^{'} \times \phi_t) + \delta_d + \delta_t + \epsilon_{dt} \end{aligned} \tag{4.6}$$

In separate regressions, the outcome variable Y_{dt} refers to various education and employment outcome variables. For education outcomes, I examine: primary school completion, years of schooling, literacy and ability to speak English. For economic activities, the outcome variable is a dummy for employment in manufacturing or a dummy for employment in services (employment in agriculture is the omitted category). For ease of presentation of the coefficients, I group observations into into 10 age bins with five-year intervals, i.e. for cohorts aged 21-25 years (born 1975-1979), aged 26-30 years (born 1970-1974), aged 31-35 years (born 1965-1969) and so on. The omitted category refers to individuals aged above 70 years (i.e. born before 1930).

Table 4.A2 provides estimates for the β_1 and β_2 coefficients for the outcome variables

examined. The peak of the tree-cutting exercises occurred in 1950-55 (see Table 4.A1). First, consider the education outcomes reported in columns (1) to (4) of Table 4.A2. In the presence of a sharp impact of the cocoa pests on the outcomes, we would have expected the treatment variable to have no impact on the outcomes examined for cohorts born prior to 1935, and for the impact of the treatment to exert some influence in subsequent years. The impact of the CSSV pest on educational outcomes would be most severe for cohorts born in the 1940s who would have made education investment decisions in the 1950s during the most active years of the CSSV campaigns. From columns (1) to (4), there is some suggestive evidence about a temporary effect on cohorts born in the 1940s in high pest infection areas. For example, for the cohort born in 1940-44, the probability of primary school completion is 6 percentage points higher, and overall schooling attainment increases by about 0.9 years. Literacy and the ability to speak English are however lower by about 9 percentage points for the subsequent cohort born between 1945-49 in these high treatment areas[5]

Overall, however, looking across cohorts born between 1930 and 1980, there appears to be no marked impact of the CSSV pest in both high or medium infection areas. Next, columns (5) and (6) also provide results on the impacts of the CSSV pests on economic activity, specifically examining if the occurrence of the pest outbreaks resulted in a transition away from agriculture into manufacturing or services. In both high and medium intensity areas, I do not observe a marked transition away from agriculture for cohorts born in the 1940s.

4.6.3 Violations of Perfect Exogeneity

In this section, I implement an approach developed by Conley, Hansen and Rossi (2012) to examine the sensitivity of the IV results to violations of the exclusion restriction assumptions. Consider the local-to-zero procedure proposed by Conley et al. This method involves directly altering the asymptotic variance matrix, and thus adjusting the computed standard errors, by introducing a term that captures the extent to which the perfect exo-

[5]Again, these results appear to be driven largely by influential observations from the abandoned areas around Koforidua. Additional results available upon request.

geneity assumption is violated. The procedure is implemented by re-running the second stage regression of government DACF transfers on opposition vote share and exogenous covariates, but now directly including the instruments (CSSV pest intensity) as shown in the regression below:

$$GovtTransfer_d = \alpha OppVoteShare_d + \gamma X_d^{'} + \delta_r + \theta * CSSV_d + \epsilon_d \tag{4.7}$$

Recall, as in previous specifications, that the CSSV instruments are the interaction of the fraction of land under high or medium pest infection and the share of land under cocoa cultivation in 1947, i.e. the fraction $High \times CocoaShare$ and $Medium \times CocoaShare$. In the case of perfect exogeneity, we would expect the coefficients on the CSSV instrument to be unidentified. Under the assumption that the coefficient is distributed as $N(0, \delta^2)$, Conley et al. suggest that sensitivity analyses may be conducted by gradually varying the magnitude of the exogeneity error and investigating the robustness of the 2SLS results.

In the sensitivity analyses, δ is set as q percent of the reduced form impact of government DACF transfers on the CSSV pest intensity and other covariates. I vary the magnitude of the exogeneity error by increasing q in steps of 2 percent between zero and 100 percent. The reduced form estimates for $High \times CocoaShare$ and fraction $Medium \times CocoaShare$ are respectively: 0.223 (se=0.121) and 0.302 (se=0.064). For values of q below 48 percent of the reduced form impact, the 2SLS results remain robust as the coefficient for opposition vote share is still negative and statistically significant. The analysis above suggests that the results presented in Table 4.5 appear robust to moderate levels of imperfect exogeneity, further increasing confidence in the IV results.

4.7 Conclusions

This study examined the impacts of the cocoa swollen shoot disease in southern Ghana, investigating its short- and long-run impacts on local political developments. It contributes to the literature examining the formation and persistence of partisanship, and how partisanship influence the distribution of resources. More broadly, it contributes to a growing literature which investigates the historical origins of various norms and beliefs (Nunn, 2012).

The Watson Commission which investigated the 1948 riots in colonial Ghana noted the political import of these disturbances, but the consequences of these riots have not been empirically examined. Using new data on cultivated cocoa acreages and spatial variation in the incidence of the CSSV pest, I examine the impacts of the disease intensity on local politics. I find that today, individuals in districts which historically experienced a high intensity of the disease pest report stronger anti-government opinions in the Afrobarometer surveys, and are more likely to attribute success in life to individual effort than government support. I trace the historical roots of these political views by examining electoral results from the 1956 Legislative Elections in colonial Ghana. I find that districts which were severely affected were more likely to vote against Nkrumah's CPP party in the 1950s, and further against the center-left party in multiparty elections in the late 1990s.

Next, I examine the implications of this result for distributive politics in Ghana in the 1990s by examining the impact of support for the opposition party on allocated transfers from the central government. I report both OLS and 2SLS results showing that an increase in opposition vote share is associated with reductions in government transfers. Finally, I conduct various checks of the plausibility of the assumptions of the exclusion restrictions of my 2SLS strategy. I do not find any persistent impacts of the CSSV pest on other outcomes such as cocoa output, educational attainment or sector of employment. Moreover, using the approach developed by Conley, Hansen and Rossi (2012), I also document that the 2SLS results remain robust to moderate forms of violations to the assumptions of the exclusion restrictions.

TABLE 4.A1: IMPACTS ON COCOA OUTPUT (see equation (4.5))

YEAR	High * CocoaShare (β_1)	Medium * CocoaShare (β_2)
1949	-50,168***	-112.6
	(13,411)	(5,303)
1950	-52,523***	1,961
	(14,656)	(6,133)
1955	-72,257***	-3,031
	(21,780)	(8,836)
1960	-39,178	39,371
	(34,683)	(28,845)
1961	-46,259	11,738
	(37,572)	(31,862)
1962	-31,091	9,403
	(35,431)	(15,698)
1963	-43,445	6,592
	(34,866)	(18,646)
1964	-33,842	5,584
	(35,044)	(24,915)
1965	-41,628	2,483
	(28,272)	(19,942)
1978	-37,841	-7,916
	(32,983)	(9,983)
1980	-33,260	-3,457
	(33,303)	(10,002)
1985	-47,091	-17,638
	(30,454)	(12,124)
1990	-50,566	-30,780
	(29,454)	(22,063)
2003	-9,987	-108,630
	(48,169)	(92,924)
2007	-34,669	-79,621
	(47,267)	(68,912)

NOTES: Dependent variable is cocoa output (in metric tonnes). Table presents coefficients on the interaction between CSSV disease pest intensity and year (see equation (4.5)). The CSSV treatment variable is given by the fraction of district land area classified as high or medium pest intensity interacted with the fraction of cultivated cocoa area as at *1948*. Regression includes controls only for districts' geographical characteristics: longitude, latitude, slope and land area. Robust standard errors are in parentheses, and clustered at district level. *** $p<0.01$, ** $p<0.05$, * $p<0.1$

TABLE 4.A2: IMPACT ON EDUCATION AND EMPLOYMENT ACTIVITY

	(1)	(2)	(3)	(4)	(5)	(6)
VARIABLES	Primary school	Years of schooling	Literacy	Able to speak English	Employed in manuf.	Employed in services
High*CocoaShare*Born (1975-79)	0.0101	0.205	-0.0542	-0.0363	-0.0279	0.0628**
	(0.0405)	(0.504)	(0.0516)	(0.0477)	(0.0340)	(0.0303)
High*CocoaShare*Born (1970-74)	0.0476	0.535	-0.0267	-0.0136	-0.0289	0.0657
	(0.0424)	(0.523)	(0.0454)	(0.0453)	(0.0290)	(0.0472)
High*CocoaShare*Born (1965-69)	0.0338	0.324	-0.0303	-0.0194	-0.0239	0.107
	(0.0404)	(0.529)	(0.0409)	(0.0390)	(0.0231)	(0.0738)
High*CocoaShare*Born (1960-64)	0.0118	0.176	-0.0442	-0.0354	-0.0227	0.116
	(0.0383)	(0.519)	(0.0359)	(0.0363)	(0.0225)	(0.0764)
High*CocoaShare*Born (1955-59)	0.0278	0.445	-0.0181	-0.0107	-0.0389*	0.0737
	(0.0413)	(0.539)	(0.0477)	(0.0470)	(0.0209)	(0.0808)
High*CocoaShare*Born (1950-54)	0.0504	0.597	-0.0107	0.00544	-0.0268	0.0626
	(0.0399)	(0.545)	(0.0315)	(0.0328)	(0.0205)	(0.0761)
High*CocoaShare*Born (1945-49)	-0.0531	-0.579	-0.0886***	-0.0885***	0.0129	0.0451
	(0.0339)	(0.443)	(0.0317)	(0.0321)	(0.0181)	(0.0714)
High*CocoaShare*Born (1940-44)	0.0585*	0.918**	-0.0151	0.000453	-0.0117	0.0793
	(0.0306)	(0.409)	(0.0313)	(0.0291)	(0.0178)	(0.0638)
High*CocoaShare*Born (1935-39)	0.0269	0.184	-0.0309	-0.0180	0.0217	0.0280
	(0.0243)	(0.331)	(0.0322)	(0.0321)	(0.0167)	(0.0419)
High*CocoaShare*Born (1930-34)	0.0215	0.240	-0.0264	0.00102	-0.0180*	0.0340
	(0.0242)	(0.339)	(0.0213)	(0.0231)	(0.0108)	(0.0286)
Medium*CocoaShare*Born (1975-79)	0.0324	0.309	-0.101*	-0.0764	-0.0658	0.0623***
	(0.0341)	(0.391)	(0.0597)	(0.0511)	(0.0440)	(0.0213)
Medium*CocoaShare*Born (1970-74)	0.0625	0.678	-0.0774	-0.0475	-0.0509	0.125***
	(0.0379)	(0.451)	(0.0569)	(0.0496)	(0.0366)	(0.0411)
Medium*CocoaShare*Born (1965-69)	0.0197	0.0675	-0.0891*	-0.0689	-0.0402	0.128**
	(0.0390)	(0.496)	(0.0512)	(0.0467)	(0.0317)	(0.0632)
Medium*CocoaShare*Born (1960-64)	0.0453	0.439	-0.0671	-0.0440	-0.0365	0.138*
	(0.0403)	(0.510)	(0.0555)	(0.0494)	(0.0306)	(0.0695)
Medium*CocoaShare*Born (1955-59)	0.0246	-0.0206	-0.0650	-0.0375	-0.0315	0.114
	(0.0384)	(0.525)	(0.0543)	(0.0484)	(0.0293)	(0.0767)
Medium*CocoaShare*Born (1950-54)	0.0119	0.139	-0.0790*	-0.0629	-0.0305	0.0966
	(0.0403)	(0.552)	(0.0456)	(0.0392)	(0.0249)	(0.0834)
Medium*CocoaShare*Born (1945-49)	0.0170	0.119	-0.0358	-0.0119	-0.00853	0.102
	(0.0414)	(0.508)	(0.0423)	(0.0385)	(0.0222)	(0.0776)
Medium*CocoaShare*Born (1940-44)	0.0330	0.242	-0.0347	-0.00676	-0.0332	0.0874
	(0.0291)	(0.368)	(0.0469)	(0.0421)	(0.0228)	(0.0655)
Medium*CocoaShare*Born (1935-39)	0.00316	0.00555	-0.0812**	-0.0799**	-0.0319	0.0661*
	(0.0305)	(0.387)	(0.0372)	(0.0379)	(0.0221)	(0.0334)
Medium*CocoaShare*Born (1930-34)	0.0196	0.212	-0.0392	-0.0381	-0.0359**	0.0288
	(0.0264)	(0.331)	(0.0256)	(0.0242)	(0.0154)	(0.0207)
Observations	567,336	567,336	567,336	567,336	567,336	567,336
R-squared	0.128	0.111	0.133	0.128	0.039	0.110

NOTES: Table presents coefficients on the interaction between CSSV disease pest intensity and age cohort bins (see equation (4.6)). The CSSV treatment variable is given by the fraction of district land area classified as high or medium pest intensity interacted with the fraction of cultivated cocoa area as at *1948*. Regression includes controls only for districts' geographical characteristics: longitude, latitude, slope, land area and cocoa soil suitability. Robust standard errors are in parentheses, and clustered at district level. *** $p<0.01$, ** $p<0.05$, * $p<0.1$

Chapter 5

Appendices

5.1 Appendix A: Data Description for Chapter 2

Treatment Intensity. I calculate treatment intensity using village gazetteers from the Tanzania 1978 Census. The treatment measure for each district is calculated as the faction of the population living in developmental villages i.e. the ratio $\frac{P_{d,r}}{P_{d,t}}$ where $P_{d,r}$ refers to population in all registered villages in district, d and $P_{d,t}$ is the total population in district, d. Note that some parts of Tanzania with already high density of nucleated settlements experienced limited resettlements such as in the Kilimanjaro region, Bukoba district and Kyela district.

The 1978 Tanzania National Census involved new and major cartographic work due to the widespread relocations in the countryside since the previous census from 1967. Useful background notes on data compilation for the census are available by Thomas (1982) *Population Data and Development Planning in Tanzania: Tanzania 1978 Population Census User Notes* and related documents available in the Africa Studies Collection at Boston University. The 1978 Census also provides additional data on scattered, migratory, and institutional populations following the re-settlement programs.

There was a debate among planners in the 1970s to estimate the precise number of persons re-settled (Thomas (1982); Maro and Mlay (1982); McCall (1985)). For example, Maro and Mlay (1982) estimate that re-settlements involved about 8 km of movement to newly cleared and communally usable land. Precise cartographic records are now available to enable estimation. Detailed maps covering the entire land surface of Tanzania (at the level of 1:100,000, prior to and following re-settlement) and available at the Harvard Map Library enable the estimation of the entire populations re-located. However, in this paper, the *treatment* refers to the experience of living in a formally registered or developmental village during the 1970s. I therefore use the share of the district population enumerated as being in registered villages by 1978.

It is important to control for the share of the urban population in each district *prior to the villagization re-settlements* so that the estimated effects of treatment do not simply reflect the degree of rural character of a given district. Prior to 1970, less than 6 percent of the total population lived in towns with a population of more than 5000 persons (Moore, 1973). Following the re-settlements of the 1970s, the 1978 census adopted very broad criteria in the definition of an urban area (see Tanzania 1978 Census Volume 1, pp 22) based on the judgment of the local census committee. For example, wards were demarcated as urban if they were judged by census officials as having urban characteristics such as the presence of a school, dispensary or market, or as being close to urban centers. There is therefore some flexibility in defining what constitutes an urban district (see Thomas, 1982 Note 5; McCall, 1985). This definition of urban areas adopted in the census differs from the official government classification as stipulated by the Tanzanian Local Government (Urban Authorities) Act of 1982 (see Thomas, 1982; Muzzini and Lindeboom, 2008).

To provide a clear and consistent definition of urban localities across various districts, I follow Thomas (1982, Note 5) and the Tanzanian Local Government (Urban Authorities) Act of 1982, in defining urban localities as cities, regional capitals (municipalities) and other townships (or town councils). To obtain a measure of urbanization prior to formation of villages, I calculate this urban share using returns for the 1967 Census (Volumes I and II). The regional capitals listed are: Arusha, Bukoba, DSM, Dodoma, Iringa, Kigoma, Lindi, Mbeya, Morogoro, Moshi, Mtwara, Musoma, Mwanza, Tabora, and Tanga; while the former townships are: Bagamoyo, Chunya, Kahama, Kilosa, Kimamba, Kondoa, Korogwe, Lushoto, Mkoani, Mpwapwa, Mwadui, Nachingwea, Nansio, Pangani, Shinyanga, Singida, Songea, and Tukuyu. Nearly three-quarters of districts have no urban populations, and government administration is often concentrated in the single major city in each region (the regional capital). Thus, to compare districts in various parts of the country, I calculate the regional urbanization rate (the share of the population in the regional capital divided by the total regional population) which is common to all districts in the region.

District Mapping. The analyses presented in this project are based on 1978 district boundaries which were in use at the time of village formation. Pre-treatment and outcomes data are obtained from various time periods, namely: 1967, 1978, 1988, 2002 and 2009. Specifically, the number of districts in mainland Tanzania increased over time as follows: 74 (in 1967), 95 (in 1978), 103 (in 1988), 119 (in 2002) and 123 (in 2009). By superimposing maps of district boundaries from the various time periods, I am able to assign contemporary districts to their historical districts. The mapping is summarized in Table A.1 below.

Education. Microdata from the 1988 Tanzania National Census provides data on educational attainment. I utilize microdata from the Integrated Public Use Micro Sample (IPUMS). The 10 percent microsample for Tanzania provides a dataset with about 2.4 million observations. Each observation lists an individual's level of educational attainment (primary, secondary, university), total years of scholing completed, and whether or not they are literate (defined as being able to read *and* write). The data is based on district administrative boundaries for 1988 which I subsequently map on to 1978 boundaries.

Occupation Outcomes. The occupational outcomes are obtained form the IPUMS microdata from the 2002 Tanzania National Census with about 3.7 million observations. The data is based on district administrative boundaries for 2002 which I subsequently map on to 1978 boundaries. The industry classifications are roughly based on the International Standard Industrial Classification (ISIC). I group these into five main categories with the following sub-divisions *agriculture* (agriculture, fishing and forestry), *manufacturing* (mining, manufacturing, electricity, gas and water), *construction* (construction services), *trading* (wholesale and retail trade), and *public administration and other services* (public administration, defense and other services, not specified).

Pre-treatment characteristics. Pre-treatment district characteristics are obtained from Jensen and Mkama (1968). The authors served in the UN Team in Physical Planning in the Tanzanian Ministry of Lands, Settlement and Water Development. They compile district-level data based on various primary sources available in Tanzania in the 1960s (see Jensen and Mkama, 1968, pp 81). Specifically, they document district-level data on the following:

population, employment sectors (agriculture, manufacturing and mining, services), industrial establishments, marketed agricultural production, livestock holdings (cattle, sheep, goats, donkeys and pigs), cooperative membership size, health facilities (number of beds in hospitals and dispensaries), education enrollment (primary school enrollment rates), trading activities, local government revenues and estimates of district level GDP.

Consumption. Per capita household consumption data is obtained from the Tanzania Household Budget Survey (THBS) 2000. THBS is a nationally representative survey of 22,178 households conducted by the Tanzanian Statistical Service. The consumption aggregate used in the analyses refers to the total consumption per adult equivalent (standardised for a 28-day period), and stated in nominal Tanzanian Shillings. This is the standard measure used in the poverty analysis in Tanzania. The 2000 HBS provides extensive data covering a large sample of 22,178 households, compared with 10,752 households covered in the 2007 Household Budget Survey.

Rainfall Data. Rainfall data is obtained from the Tanzania Meteorological Agency (TMA) in Dar es Salaam. The network of rainfall stations on mainland Tanzania was maintained by the East Africa Meteorological Department for the period 1930-1960, and subsequently by the Tanzania Meteorological Agency (see Johnson, 1962; Nieuwolt, 1977). Data is provided from three main types of stations: synoptic stations, agrometeorological stations and rain gauge stations. There are two main agroclimatic areas in Tanzania — with unimodal and bimodal rainfall areas. Bimodal rainfall areas have two planting seasons: the minor short-rains cropping season (or *vuli*) and the long-rains season (or *masika*). For both areas, the main period of agricultural activity occurs during the *masika* season (or long-rains) with peak rainfall around March to June each year. I examine the *masika* season rainfall in this project.

I analyze monthly rainfall data for 108 rainfall stations covering most districts (based on 2009 administrative boundaries) for the period 1960-2010. Below, I document a few exceptions. For the rainfall dataset obtained from TMA, Meatu district is the only district without a rainfall gauge or station. I replace this district with data from the nearby Itinje Dispensary on the border between Meatu and Maswa districts, both in Shinyanga Region. Based on 2009 administrative boundaries, the following three districts: Nkasi, Sumbawanga Urban and Sumbawanga Rural all comprised Sumbawanga in 1978. There is only one reliable station in Sumbawanga Urban, and this is used for the entire district. Rainfall data is lacking in Tabora Rural which completely surrounds Tabora Urban. I therefore assign data for the Tabora Urban synoptic station to Tabora Rural (as defined in 1978). Today, Tabora Rural is comprised of Sikonge and Uyui. There is no continuous rainfall data for Igunga district, so I replace this with data for the neighboring Nzega district. Continuous rainfall data is available for Kiteto district, but missing only for 1976. Thus for 1976, I replace Kiteto data with the average of two nearby stations, namely Kondoa Maji and Kongwa PRS.

Electoral Data. Electoral results for the 2005 Presidential Elections are obtained from the National Electoral Commission of Tanzania (results also published online). I also compile election results from the 1970 Tanzanian National Election to assess variation in support for Nyerere's TANU Party across districts, prior to village formation. The results of the 1970 elections are publised by the Election Study Committee (1974), and provides district level data on fraction of eligible voters who registered to vote; voter turnout; and fraction of population voting *yes* in support of Nyerere in the one-party presidential election.

Local Government Tax Revenues. Local government budget summaries is obtained from PMO-RALG (Prime Ministers Office Regional Administration and Local Government, *TAMISEMI*) in Dodoma, Tanzania. I use data beginning in 2005/6 financial year. This is the earliest year for which data is available from PMO-RALG. This data is available online at the Tanzanian Government website: http://www.logintanzania.net/ (downloaded as at 1 January 2012). I use cumulative totals at the end of the fourth quarter of each fiscal year. The agricultural taxes (produce cess) is in units of Tanzanian shillings per capita, and normalized by the fraction of the population employed in agriculture, based on 2002 Tanzanian Agricultural Census.

Geographic controls. Data on geographic characteristics is obtained by overlaying the 1978 district administrative map of Tanzania on elevation and slope data obtained from SRTM/ArcMaps and computing the mean area-weighted altitude and slope. I also compute the latitude and longitude coordinates at the centroid of each district, and the distance from the centroid to the coast and to the nearest railway line (in 1975).

Community Participation and Public Goods Provision. Data on community participation and public goods provision is obtained from the Tanznaian National Household Panel Survey of 2008/9. The household survey is comprised of a nationally representative sample of 3,280 households and 16,710 individuals and conducted by the Tanzanian National Bureau of Statistics. The survey instruments include questions on respondents' districts of birth, enabling me to link migrants to their initial districts of origin.

Ethnolinguistic Fragmentation. I calculate ethnolinguistic fragmentation, ELF_d as: $(1 - \sum e_{id}^2)$, where e_i is the fraction of the population in district, d, belonging to ethnic group, i. ELF_d measures the probability that any two randomly chosen individuals in district, d, belong to different ethnic group (Mauro, 1995; Posner 2004). A high ELF_d implies a more ethnically diverse polity. In this project, I compute ELF_d based on the 1957 Census, which is the most recent census with data on district-level ethnic populations.

Afrobarometer Data. For outcomes on political participation, and perceptions on corruption, democracy, and nationalism, I utilize individual level responses from the Afrobarometer Surveys for Tanzania. The Afrobarometer survey is nationally representative, and based on interviews with citizens of voting age in Tanzania. I base the analyses in this paper on three rounds of surveys conducted in 2001 (Round 1, n = 2,198), 2005 (Round 3, n = 1,304), and 2008 (Round 4, n = 1,208). I am unable to use Round 2 of the Tanzania surveys as the publicly available dataset lacks district identifiers.

5.2 Appendix B: Additional Tables

Table A1: Mapping of Districts

REGION NAME (2009)	DISTRICT NAME 2009	2002	1978	1967
Dodoma	Kondoa	Kondoa	Kondoa	Kondoa
Dodoma	Mpwapwa	Mpwapwa	Mpwapwa	Mpwapwa
Dodoma	Kongwa	Kongwa	Mpwapwa	Mpwapwa
Dodoma	Dodoma Rural	Dodoma Rural	Dodoma Rural	Dodoma
Dodoma	Dodoma Urban	Dodoma Urban	Dodoma Urban	Dodoma
Dodoma	Bahi	Dodoma Rural	Dodoma Rural	Dodoma
Arusha	Monduli	Monduli	Monduli	Masai
Arusha	Arumeru	Arumeru	Arumeru	Arusha
Arusha	Arusha	Arusha	Arusha	Arusha
Arusha	Karatu	Karatu	Mbulu	Mbulu
Arusha	Ngorongoro	Ngorongoro	Monduli	Masai
Kilimanjaro	Rombo	Rombo	Rombo	Kilimanjaro
Kilimanjaro	Mwanga	Mwanga	Pare	Pare
Kilimanjaro	Same	Same	Pare	Pare
Kilimanjaro	Moshi Rural	Moshi Rural	Moshi Rural	Kilimanjaro
Kilimanjaro	Hai	Hai	Hai	Kilimanjaro
Kilimanjaro	Moshi Urban	Moshi Urban	Moshi Urban	Kilimanjaro

Table A1: Mapping of Districts (contd.)

REGION NAME (2009)	DISTRICT NAME 2009	2002	1978	1967
Tanga	Lushoto	Lushoto	Lushoto	Lushoto
Tanga	Korogwe	Korogwe	Korogwe	Korogwe
Tanga	Muheza	Muheza	Muheza	Tanga
Tanga	Tanga	Tanga	Tanga	Tanga
Tanga	Pangani	Pangani	Pangani	Pangani
Tanga	Handeni	Handeni	Handeni	Handeni
Tanga	Kilindi	Kilindi	Handeni	Handeni
Tanga	Mkinga	Muheza	Muheza	Tanga
Morogoro	Kilosa	Kilosa	Kilosa	Kilosa
Morogoro	Morogoro	Morogoro	Morogoro Rural	Morogoro
Morogoro	Kilombero	Kilombero	Kilombero	Ulanga
Morogoro	Ulanga	Ulanga	Mahenge	Ulanga
Morogoro	Morogoro Urban	Morogoro Urban	Morogoro Urban	Morogoro
Morogoro	Mvomero	Mvomero	Morogoro Rural	Morogoro
Pwani	Bagamoyo	Bagamoyo	Bagamoyo	Bagamoyo
Pwani	Kibaha	Kibaha	Kisarawe	Kisarawe
Pwani	Kisarawe	Kisarawe	Kisarawe	Kisarawe
Pwani	Mkuranga	Mkuranga	Kisarawe	Kisarawe
Pwani	Rufiji	Rufiji	Rufiji	Rufiji
Pwani	Mafia	Mafia	Mafia	Mafia

Table A1: Mapping of Districts (contd.)

REGION NAME (2009)	DISTRICT NAME 2009	2002	1978	1967
Dar es Salaam	Kinondoni	Kinondoni	Kinondoni	Mzizima
Dar es salaam	Ilala	Ilala	Ilala	Mzizima
Dar es Salaam	Temeke	Temeke	Temeke	Mzizima
Lindi	Kilwa	Kilwa	Kilwa	Kilwa
Lindi	Lindi Rural	Lindi Rural	Lindi Rural	Lindi
Lindi	Nachingwea	Nachingwea	Nachingwea	Nachingwea
Lindi	Liwale	Liwale	Liwale	Nachingwea
Lindi	Ruangwa	Ruangwa	Lindi Rural	Lindi
Lindi	Lindi Urban	Lindi Urban	Lindi Urban	Lindi
Mtwara	Mtwara Rural	Mtwara Rural	Mtwara Rural	Mtwara
Mtwara	Newala	Newala	Newala	Newala
Mtwara	Masasi	Masasi	Masasi	Masasi
Mtwara	Tandahimba	Tandahimba	Newala	Newala
Mtwara	Mtwara Urban	Mtwara Urban	Mtwara Urban	Mtwara
Ruvuma	Tunduru	Tunduru	Tunduru	Tunduru
Ruvuma	Songea Rural	Songea Rural	Songea Rural	Songea
Ruvuma	Mbinga	Mbinga	Mbinga	Mbinga
Ruvuma	Songea Urban	Songea Urban	Songea Urban	Songea
Ruvuma	Namtumbo	Namtumbo	Songea Rural	Songea
Iringa	Iringa Rural	Iringa Rural	Iringa Rural	Iringa

Table A1: Mapping of Districts (contd.)

REGION NAME (2009)	DISTRICT NAME 2009	2002	1978	1967
Iringa	Mufindi	Mufindi	Mufindi	Mufindi
Iringa	Makete	Makete	Njombe	Njombe
Iringa	Njombe	Njombe	Njombe	Njombe
Iringa	Ludewa	Ludewa	Ludewa	Njombe
Iringa	Iringa Urban	Iringa Urban	Iringa Urban	Iringa
Iringa	Kilolo	Kilolo	Iringa Rural	Iringa
Mbeya	Chunya	Chunya	Chunya	Chunya
Mbeya	Mbeya Rural	Mbeya (R)	Mbeya Rural	Mbeya
Mbeya	Kyela	Kyela	Kyela	Rungwe
Mbeya	Rungwe	Rungwe	Rungwe	Rungwe
Mbeya	Ileje	Ileje	Ileje	Rungwe
Mbeya	Mbozi	Mbozi	Mbozi	Mbozi
Mbeya	Mbarali	Mbarali	Mbeya Rural	Mbeya
Mbeya	Mbeya Urban	Mbeya Urban	Mbeya Urban	Mbeya
Singida	Iramba	Iramba	Iramba	Iramba
Singida	Singida Rural	Singida Rural	Singida Rural	Singida
Singida	Manyoni	Manyoni	Manyoni	Manyoni
Singida	Singida Urban	Singida Urban	Singida Urban	Singida
Tabora	Nzega	Nzega	Nzega	Nzega

Table A1: Mapping of Districts (contd.)

REGION NAME (2009)	**DISTRICT NAME 2009**	**2002**	**1978**	**1967**
Tabora	Igunga	Igunga	Igunga	Nzega
Tabora	Uyui	Uyui	Tabora Rural	Tabora
Tabora	Urambo	Urambo	Urambo	Tabora
Tabora	Sikonge	Sikonge	Tabora Rural	Tabora
Tabora	Tabora Urban	Tabora Urban	Tabora Urban	Tabora
Rukwa	Mpanda	Mpanda	Mpanda	Mpanda
Rukwa	Sumbawanga Rural	Sumbawanga Rural	Sumbawanga Rural	Sumbawanga
Rukwa	Nkasi	Nkasi	Sumbawanga Rural	Sumbawanga
Rukwa	Sumbawanga Urban	Sumbawanga Urban	Sumbawanga Urban	Sumbawanga
Kigoma	Kibondo	Kibondo	Kibondo	Kibondo
Kigoma	Kasulu	Kasulu	Kasulu	Kasulu
Kigoma	Kigoma Rural	Kigoma Rural	Kigoma Rural	Kigoma
Kigoma	Kigoma Urban	Kigoma Urban	Kigoma Urban	Kigoma
Shinyanga	Bariadi	Bariadi	Bariadi	Maswa
Shinyanga	Maswa	Maswa	Maswa	Maswa
Shinyanga	Shinyanga Rural	Shinyanga Rural	Shinyanga Rural	Shinyanga
Shinyanga	Kahama	Kahama	Kahama	Kahama
Shinyanga	Bukombe	Bukombe	Kahama	Kahama
Shinyanga	Meatu	Meatu	Maswa	Maswa
Shinyanga	Shinyanga Urban	Shinyanga Urban	Shinyanga Urban	Shinyanga

Table A1: Mapping of Districts (contd.)

REGION NAME (2009)	**DISTRICT NAME 2009**	**2002**	**1978**	**1967**
Shinyanga	Kishapu	Kishapu	Shinyanga Rural	Shinyanga
Kagera	Karagwe	Karagwe	Karagwe	Karagwe
Kagera	Bukoba Rural	Bukoba Rural	Bukoba Rural	Bukoba
Kagera	Muleba	Muleba	Muleba	Bukoba
Kagera	Biharamulo	Biharamulo	Biharamulo	Biharamulo
Kagera	Ngara	Ngara	Ngara	Ngara
Kagera	Bukoba Urban	Bukoba Urban	Bukoba Urban	Bukoba
Kagera	Chato	Biharamulo	Biharamulo	Biharamulo
Kagera	Misenye	Bukoba Rural	Bukoba Rural	Bukoba
Mwanza	Ukerewe	Ukerewe	Ukerewe	Ukerewe
Mwanza	Magu	Magu	Magu	Mwanza
Mwanza	Nyamagana	Nyamagana	Mwanza	Mwanza
Mwanza	Kwimba	Kwimba	Kwimba	Kwimba
Mwanza	Sengerema	Sengerema	Sengerema	Ukerewe
Mwanza	Geita	Geita	Geita	Geita
Mwanza	Missungwi	Missungwi	Kwimba	Kwimba
Mwanza	Ilemela	Ilemela	Mwanza	Mwanza
Mara	Tarime	Tarime	Tarime	Mara
Mara	Serengeti	Serengeti	Serengeti	Musoma
Mara	Musoma Rural	Musoma Rural	Musoma Rural	Musoma

Table A1: Mapping of Districts (contd.)

REGION NAME (2009)	**DISTRICT NAME** 2009	2002	1978	1967
Mara	Bunda	Bunda	Musoma Rural	Musoma
Mara	Musoma Urban	Musoma Urban	Musoma Urban	Musoma
Manyara	Babati	Babati	Hanang	Mbulu
Manyara	Hanang	Hanang	Hanang	Mbulu
Manyara	Mbulu	Mbulu	Mbulu	Mbulu
Manyara	Simanjiro	Simanjiro	Kiteto	Masai
Manyara	Kiteto	Kiteto	Kiteto	Masai
kaskazini	Kaskazini ’A’	Kaskazini ’A’	Zanzibar	Zanzibar
kaskazini	Kaskazini ’B’	Kaskazini ’B’	Zanzibar	Zanzibar
Kusini	Kati	Kati	Zanzibar	Zanzibar
Kusini	Kusini	Kusini	Zanzibar	Zanzibar
Mjini Magharibi	Magharibi	Magharibi	Zanzibar	Zanzibar
Mjini Magharibi	Mjini	Mjini	Zanzibar	Zanzibar
Kaskazini Pemba	Wete	Wete	Pemba	Pemba
Kaskazini Pemba	Micheweni	Micheweni	Pemba	Pemba
Kusini Pemba	Chake	Chake	Pemba	Pemba
Kusini Pemba	Mkoani	Mkoani	Pemba	Pemba

TABLE B.1: COEFFICIENTS OF INTERACTIONS BETWEEN DUMMIES INDICATING THE AGE IN 1974 AND FRACTION OF DISTRICT POPULATION LIVING IN DEVELOPMENT VILLAGES

	Dep. Variable: Primary School Completion			
Age in 1974	(1)	(2)	(3)	(4)
1	0.142***	0.142***	0.133***	0.102***
	(0.0361)	(0.0361)	(0.0370)	(0.0297)
2	0.160***	0.160***	0.151***	0.123***
	(0.0336)	(0.0336)	(0.0343)	(0.0251)
3	0.168***	0.168***	0.159***	0.129***
	(0.0324)	(0.0324)	(0.0337)	(0.0244)
4	0.157***	0.157***	0.147***	0.117***
	(0.0331)	(0.0331)	(0.0345)	(0.0263)
5	0.158***	0.158***	0.149***	0.119***
	(0.0312)	(0.0312)	(0.0329)	(0.0253)
6	0.128***	0.127***	0.118***	0.0878***
	(0.0325)	(0.0325)	(0.0345)	(0.0276)
7	0.158***	0.158***	0.158***	0.154***
	(0.0297)	(0.0297)	(0.0298)	(0.0299)
8	0.142***	0.142***	0.143***	0.138***
	(0.0316)	(0.0316)	(0.0317)	(0.0317)
9	0.130***	0.130***	0.130***	0.126***
	(0.0301)	(0.0301)	(0.0302)	(0.0305)
10	0.116***	0.116***	0.116***	0.111***
	(0.0301)	(0.0301)	(0.0302)	(0.0304)
11	0.0968***	0.0968***	0.0973***	0.0935***
	(0.0277)	(0.0276)	(0.0277)	(0.0276)
12	0.0801***	0.0801***	0.0798***	0.0743**
	(0.0287)	(0.0286)	(0.0287)	(0.0287)
13	0.0744***	0.0744***	0.0745***	0.0691**
	(0.0268)	(0.0268)	(0.0269)	(0.0268)
14	0.0482*	0.0482*	0.0477*	0.0431
	(0.0274)	(0.0274)	(0.0274)	(0.0273)
15	0.0303	0.0303	0.0299	0.0269
	(0.0245)	(0.0245)	(0.0246)	(0.0246)
16	0.0163	0.0163	0.0166	0.0154
	(0.0274)	(0.0274)	(0.0274)	(0.0276)
17	0.00629	0.00624	0.00615	0.00258
	(0.0227)	(0.0227)	(0.0227)	(0.0227)
18	-0.00298	-0.00303	-0.00336	-0.00296
	(0.0203)	(0.0203)	(0.0203)	(0.0206)
19	0.00233	0.00226	0.00203	0.00282
	(0.0237)	(0.0237)	(0.0238)	(0.0239)
20	-0.000859	-0.000906	-0.00133	-0.00266
	(0.0203)	(0.0203)	(0.0203)	(0.0205)
21	0.0205	0.0205	0.0213	0.0242
	(0.0217)	(0.0217)	(0.0216)	(0.0219)
22	0.00742	0.00734	0.00754	0.00690
	(0.0241)	(0.0241)	(0.0241)	(0.0243)
23	-0.0122	-0.0123	-0.0117	-0.0104
	(0.0221)	(0.0221)	(0.0220)	(0.0222)
24	0.00204	0.00201	0.00222	0.00471
	(0.0187)	(0.0187)	(0.0187)	(0.0190)
Observations	827,027	827,027	820,852	798,324
R-squared	0.155	0.155	0.155	0.155

*** p<0.01, ** p<0.05, * p<0.1. Dependent variable is dummy for primary school completion. Data is obtained from IPUMS microsample of 1988 Tanzania Census. All columns include fixed effects for year of birth cohorts and 1978 districts. Controls enter the regression interacted with the young dummy. Column 2 controls for pre-treatment primary enrollment in 1970. Column 3 includes controls for health infrastructure in 1970. Column 4 includes various geographic and demographic controls. Robust standard errors are in parentheses and clustered at the 1978 district boundaries.

TABLE C.1: COHORT EFFECTS FOR POLITICAL PARTICIPATION

	(1)	(2)	(3)	(4)	(5)	(6)
PANEL A						
Dep. Var: Contacted local government official in the past year?						
Born 1980-1989	0.149*	0.168*	0.137	0.148	0.139	0.138
	(0.0762)	(0.0856)	(0.101)	(0.101)	(0.103)	(0.104)
Born 1970-1979	0.222***	0.219***	0.270**	0.288***	0.285***	0.285***
	(0.0747)	(0.0815)	(0.108)	(0.106)	(0.108)	(0.108)
Born 1960-1969	0.211***	0.208**	0.229**	0.257**	0.252**	0.251**
	(0.0725)	(0.0858)	(0.105)	(0.101)	(0.103)	(0.103)
Born 1950-1959	0.206**	0.171	0.163	0.188	0.181	0.186
	(0.103)	(0.111)	(0.134)	(0.135)	(0.136)	(0.136)
Observations	4,283	4,283	4,283	4,189	4,189	4,189
R-squared	0.119	0.122	0.125	0.129	0.129	0.129
PANEL B						
Dep. Var: Contacted any government civil service official in the past year?						
Born 1980-1989	0.0815	0.125*	0.133*	0.153*	0.141*	0.138*
	(0.0678)	(0.0735)	(0.0782)	(0.0821)	(0.0808)	(0.0773)
Born 1970-1979	0.0915*	0.110*	0.129**	0.135**	0.124*	0.119*
	(0.0489)	(0.0577)	(0.0645)	(0.0663)	(0.0661)	(0.0620)
Born 1960-1969	0.0739	0.0975	0.106	0.127	0.111	0.105
	(0.0671)	(0.0755)	(0.0788)	(0.0825)	(0.0783)	(0.0745)
Born 1950-1959	0.0115	0.0201	0.0540	0.0800	0.0578	0.0666
	(0.0922)	(0.0948)	(0.0983)	(0.102)	(0.0979)	(0.0960)
Observations	4,278	4,278	4,278	4,184	4,184	4,184
R-squared	0.079	0.082	0.084	0.085	0.087	0.089
PANEL C						
Dep. Var: Able to correctly name local MP?						
Born 1980-1989	0.273***	0.309***	0.225*	0.195	0.218*	0.193
	(0.0989)	(0.103)	(0.114)	(0.120)	(0.120)	(0.123)
Born 1970-1979	0.144	0.143	0.100	0.0885	0.119	0.128
	(0.106)	(0.111)	(0.112)	(0.113)	(0.114)	(0.117)
Born 1960-1969	0.150	0.167	0.0704	0.0475	0.0832	0.103
	(0.168)	(0.156)	(0.156)	(0.150)	(0.147)	(0.148)
Born 1950-1959	0.0335	-0.0133	-0.0536	-0.0943	-0.0718	-0.0574
	(0.160)	(0.160)	(0.181)	(0.184)	(0.182)	(0.180)
Observations	2,277	2,277	2,277	2,246	2,246	2,246
R-squared	0.205	0.209	0.215	0.219	0.222	0.226
Geographic Controls	NO	YES	YES	YES	YES	YES
Demographic Controls	NO	NO	YES	YES	YES	YES
Agricultural Controls	NO	NO	NO	YES	YES	YES
Education Controls	NO	NO	NO	NO	YES	YES
Local Govt Controls	NO	NO	NO	NO	NO	YES

NOTES: Data is obtained from the Afrobarometer Tanzania Surveys: Rounds 3 and 4 data are used in Panels A and B; and Round 1 data for Panel C. Results are presented for coefficients on four birth cohorts above compared to the omitted category which was born prior to 1950. Geographic controls are for latitude, mean weighted slope and altitude. Demographic controls are for mean pre-treatment ethnolinguistic fragmentation, fraction of moslems and fraction of christians in the population based on the 1967 Tanzania Census. Agricultural controls are for per capita populations of cattle, sheep, and goats (Jensen and Mkama, 1968). Education control is for school enrollment rate in 1967 and local government controls are for per capita local government revenues in 1968-1970 (Jensen and Mkama, 1968). Robust standard errors are in parentheses and clustered at the 1978 district boundaries. *** $p<0.01$, ** $p<0.05$, * $p<0.1$.

TABLE C.2: COHORT EFFECTS ON ONE-PARTY RULE AND DEMOCRACY

	(1)	(2)	(3)	(4)	(5)	(6)
PANEL A *Dep. Var: Dummy for supporting one-party rule*						
Born 1980-1989	-0.138	-0.179	-0.316*	-0.283*	-0.297*	-0.287*
	(0.137)	(0.148)	(0.173)	(0.161)	(0.158)	(0.159)
Born 1970-1979	-0.0709	-0.144	-0.319**	-0.268*	-0.277*	-0.280*
	(0.121)	(0.139)	(0.154)	(0.150)	(0.150)	(0.150)
Born 1960-1969	-0.140	-0.216	-0.258	-0.266*	-0.277*	-0.296*
	(0.133)	(0.151)	(0.158)	(0.158)	(0.157)	(0.162)
Born 1950-1959	-0.176	-0.225*	-0.299**	-0.275*	-0.276*	-0.296*
	(0.115)	(0.127)	(0.149)	(0.150)	(0.153)	(0.150)
Observations	2,014	2,014	2,014	1,950	1,950	1,950
R-squared	0.163	0.172	0.179	0.180	0.182	0.185
PANEL B *Dep. Var: Dummy for supporting democracy*						
Born 1980-1989	0.0164	0.144	0.281	0.291	0.305*	0.320*
	(0.138)	(0.144)	(0.178)	(0.177)	(0.179)	(0.181)
Born 1970-1979	-0.0820	0.0176	0.136	0.116	0.123	0.139
	(0.109)	(0.118)	(0.140)	(0.142)	(0.147)	(0.144)
Born 1960-1969	-0.0227	0.133	0.118	0.137	0.150	0.173
	(0.112)	(0.123)	(0.132)	(0.139)	(0.144)	(0.145)
Born 1950-1959	-0.0191	0.0917	0.175	0.165	0.169	0.174
	(0.106)	(0.114)	(0.141)	(0.145)	(0.149)	(0.141)
Observations	2,014	2,014	2,014	1,950	1,950	1,950
R-squared	0.202	0.208	0.219	0.223	0.225	0.226
Geographic Controls	NO	YES	YES	YES	YES	YES
Demographic Controls	NO	NO	YES	YES	YES	YES
Agricultural Controls	NO	NO	NO	YES	YES	YES
Education Controls	NO	NO	NO	NO	YES	YES
Local Govt Controls	NO	NO	NO	NO	NO	YES

NOTES: Data is obtained from Round 1 of the Tanzania Afrobarometer Survey. Results are presented for coefficients on four birth cohorts above compared to the omitted category which was born prior to 1950. Geographic controls are for latitude, mean weighted slope and altitude. Demographic controls are for mean pre-treatment ethnolinguistic fragmentation, fraction of moslems and fraction of christians in the population based on the 1967 Tanzania Census. Agricultural controls are for per capita populations of cattle, sheep, and goats (Jensen and Mkama, 1968). Education control is for school enrollment rate in 1967 and local government controls are for per capita local government revenues in 1968-1970 (Jensen and Mkama, 1968). Robust standard errors are in parentheses and clustered at the 1978 district boundaries. *** $p<0.01$, ** $p<0.05$, * $p<0.1$.

TABLE C.3: COHORT EFFECTS FOR PERCEPTIONS OF CORRUPTION

	(1)	(2)	(3)	(4)	(5)	(6)
PANEL A						
Dep. Var: Dummy equal to one if corruption is prevalent among elected leaders (MPs)?						
Mean Dep. Var. = 0.531						
Born 1980-1989	0.214	0.259	0.288	0.327*	0.285	0.282
	(0.162)	(0.167)	(0.180)	(0.179)	(0.196)	(0.197)
Born 1970-1979	0.109	0.235	0.224	0.326*	0.310	0.319*
	(0.146)	(0.158)	(0.175)	(0.163)	(0.189)	(0.191)
Born 1960-1969	0.210	0.315*	0.340*	0.363*	0.312	0.294
	(0.184)	(0.183)	(0.191)	(0.200)	(0.217)	(0.223)
Born 1950-1959	0.0308	0.0785	0.151	0.203	0.188	0.205
	(0.191)	(0.190)	(0.218)	(0.231)	(0.237)	(0.234)
Observations	2,238	2,238	2,238	2,207	2,207	2,207
R-squared	0.167	0.172	0.174	0.174	0.175	0.176
PANEL B						
Dep. Var: Dummy equal to one if corruption is prevalent among government officials?						
Mean Dep. Var. = 0.597						
Born 1980-1989	0.308	0.332*	0.329	0.388*	0.365*	0.365*
	(0.187)	(0.185)	(0.198)	(0.195)	(0.201)	(0.205)
Born 1970-1979	0.126	0.157	0.0481	0.158	0.0972	0.103
	(0.148)	(0.164)	(0.163)	(0.144)	(0.163)	(0.165)
Born 1960-1969	0.127	0.161	0.202	0.228	0.174	0.156
	(0.159)	(0.160)	(0.166)	(0.167)	(0.178)	(0.184)
Born 1950-1959	0.0376	0.0739	-0.000521	0.0573	0.00804	-0.0127
	(0.148)	(0.151)	(0.172)	(0.173)	(0.180)	(0.184)
Observations	2,238	2,238	2,238	2,207	2,207	2,207
R-squared	0.138	0.141	0.146	0.147	0.148	0.149
Geographic Controls	NO	YES	YES	YES	YES	YES
Demographic Controls	NO	NO	YES	YES	YES	YES
Agricultural Controls	NO	NO	NO	YES	YES	YES
Education Controls	NO	NO	NO	NO	YES	YES
Local Govt Controls	NO	NO	NO	NO	NO	YES

NOTES: Data is obtained from the Afrobarometer Tanzania Surveys (Rounds 3 and 4). Results are presented for coefficients on four birth cohorts above compared to the omitted category which was born prior to 1950. Geographic controls are for latitude, mean weighted slope and altitude. Demographic controls are for mean pre-treatment ethnolinguistic fragmentation, fraction of moslems and fraction of christians in the population based on the 1967 Tanzania Census. Agricultural controls are for per capita populations of cattle, sheep, and goats (Jensen and Mkama, 1968). Education control is for school enrollment rate in 1967 and local government controls are for per capita local government revenues in 1968-1970 (Jensen and Mkama, 1968). Robust standard errors are in parentheses and clustered at the 1978 district boundaries. *** $p<0.01$, ** $p<0.05$, * $p<0.1$.

TABLE C.3: COHORT EFFECTS FOR PERCEPTIONS OF CORRUPTION (contd.)

	(1)	(2)	(3)	(4)	(5)	(6)
			PANEL C			
	Dep. Var: Dummy equal to one if corruption is prevalent among the police?					
	Mean Dep. Var. = 0.784					
Born 1980-1989	0.134	0.234**	0.116	0.117	0.135	0.124
	(0.122)	(0.111)	(0.124)	(0.133)	(0.148)	(0.150)
Born 1970-1979	0.105	0.252**	0.134	0.190	0.180	0.198
	(0.107)	(0.0969)	(0.117)	(0.116)	(0.132)	(0.129)
Born 1960-1969	0.167	0.265**	0.174	0.165	0.0945	0.100
	(0.122)	(0.117)	(0.126)	(0.130)	(0.148)	(0.147)
Born 1950-1959	0.0569	0.178	0.0355	0.0792	0.0532	0.0255
	(0.138)	(0.132)	(0.162)	(0.173)	(0.176)	(0.185)
Observations	2,238	2,238	2,238	2,207	2,207	2,207
R-squared	0.102	0.111	0.117	0.123	0.126	0.128
Geographic Controls	NO	YES	YES	YES	YES	YES
Demographic Controls	NO	NO	YES	YES	YES	YES
Agricultural Controls	NO	NO	NO	YES	YES	YES
Education Controls	NO	NO	NO	NO	YES	YES
Local Govt Controls	NO	NO	NO	NO	NO	YES

NOTES: Data is obtained from the Afrobarometer Tanzania Surveys (Rounds 3 and 4). Results are presented for coefficients on four birth cohorts above compared to the omitted category which was born prior to 1950. Geographic controls are for latitude, mean weighted slope and altitude. Demographic controls are for mean pre-treatment ethnolinguistic fragmentation, fraction of moslems and fraction of christians in the population based on the 1967 Tanzania Census. Agricultural controls are for per capita populations of cattle, sheep, and goats (Jensen and Mkama, 1968). Education control is for school enrollment rate in 1967 and local government controls are for per capita local government revenues in 1968-1970 (Jensen and Mkama, 1968). Robust standard errors are in parentheses and clustered at the 1978 district boundaries. *** p<0.01, ** p<0.05, * p<0.1.

TABLE C.4: COHORT EFFECTS FOR PERCEPTIONS ON ETHNIC AND NATIONAL IDENTITY

	(1)	(2)	(3)	(4)	(5)	(6)
PANEL A						
Dep. Var: Dummy equal to one if respondent prefers national to ethnic identity						
Mean Dep. Var. = 0.753						
Born 1980-1989	-0.0758	-0.162	-0.193*	-0.181	-0.192	-0.196*
	(0.104)	(0.106)	(0.114)	(0.117)	(0.116)	(0.116)
Born 1970-1979	-0.0707	-0.174*	-0.184*	-0.179	-0.185*	-0.179*
	(0.101)	(0.0934)	(0.108)	(0.109)	(0.109)	(0.103)
Born 1960-1969	-0.154	-0.190*	-0.229**	-0.214*	-0.225**	-0.222**
	(0.0991)	(0.0998)	(0.110)	(0.112)	(0.110)	(0.107)
Born 1950-1959	-0.0218	-0.0922	-0.0366	-0.0177	-0.0261	-0.0250
	(0.132)	(0.123)	(0.137)	(0.139)	(0.137)	(0.139)
Observations	4,224	4,224	4,224	4,130	4,130	4,130
R-squared	0.072	0.076	0.079	0.081	0.081	0.083
PANEL B						
Dep. Var: Dummy equal to one if respondent trusts Tanzanians of other tribes						
Mean Dep. Var. = 0.672						
Born 1980-1989	-0.0307	-0.0260	-0.0409	-0.0632	-0.0735	-0.0763
	(0.0724)	(0.0815)	(0.0967)	(0.100)	(0.0981)	(0.0977)
Born 1970-1979	0.00141	0.0526	0.0602	0.0401	0.0159	0.0201
	(0.0708)	(0.0695)	(0.0751)	(0.0756)	(0.0715)	(0.0696)
Born 1960-1969	0.0249	0.0213	0.0273	0.0238	0.0196	0.0195
	(0.0716)	(0.0761)	(0.0836)	(0.0804)	(0.0788)	(0.0803)
Born 1950-1959	-0.101	-0.125	-0.0888	-0.121	-0.128	-0.136*
	(0.0670)	(0.0765)	(0.0919)	(0.0783)	(0.0783)	(0.0783)
Observations	3,178	3,178	3,178	3,106	3,106	3,106
R-squared	0.101	0.106	0.108	0.111	0.113	0.114
Geographic Controls	NO	YES	YES	YES	YES	YES
Demographic Controls	NO	NO	YES	YES	YES	YES
Agricultural Controls	NO	NO	NO	YES	YES	YES
Education Controls	NO	NO	NO	NO	YES	YES
Local Govt Controls	NO	NO	NO	NO	NO	YES

NOTES: Data is obtained from the Tanzania Afrobarometer Survey. Panel A is based on results from Rounds 1, 3 and 4. Panel B is based on results from Rounds 1 and 3. Results are presented for coefficients on four birth cohorts above compared to the omitted category which was born prior to 1950. Geographic controls are for latitude, mean weighted slope and altitude. Demographic controls are for mean pre-treatment ethnolinguistic fragmentation, fraction of moslems and fraction of christians in the population based on the 1967 Tanzania Census. Agricultural controls are for per capita populations of cattle, sheep, and goats (Jensen and Mkama, 1968). Education control is for school enrollment rate in 1967 and local government controls are for per capita local government revenues in 1968-1970 (Jensen and Mkama, 1968). Robust standard errors are in parentheses and clustered at the 1978 district boundaries. *** p<0.01, ** p<0.05, * p<0.1.

TABLE C.5: ROBUSTNESS CHECKS

	(1) Main Equation Eqn. (2.4)	(2) Control for Voter Turnout (1970)	(3) Control for Share of Nyerere Vote (1970)	(4) Controls for Turnout and Share of Nyerere vote (1970)
Contact Local Govt Official	0.156**	0.144**	0.140**	0.139**
	(0.0633)	(0.0628)	(0.0639)	(0.0649)
Contact Civil Service	0.0887*	0.0894*	0.0880*	0.0868*
	(0.0455)	(0.0483)	(0.0483)	(0.0477)
Identify MP	0.224**	0.214**	0.211**	0.211**
	(0.0993)	(0.0994)	(0.101)	(0.101)
Support one-party state	-0.209*	-0.178	-0.183	-0.192
	(0.125)	(0.121)	(0.123)	(0.121)
Support democracy	0.190*	0.165	0.170	0.178*
	(0.107)	(0.102)	(0.103)	(0.101)
Corrupt MPs	0.227*	0.265*	0.272*	0.272*
	(0.135)	(0.147)	(0.152)	(0.153)
Corrupt civil servants	0.300**	0.334**	0.335**	0.333**
	(0.147)	(0.158)	(0.163)	(0.165)
Corrupt police	0.196*	0.229*	0.244*	0.247*
	(0.116)	(0.120)	(0.125)	(0.126)
Nation or tribe	-0.163**	-0.161**	-0.164**	-0.167**
	(0.0695)	(0.0691)	(0.0711)	(0.0726)
Trust other tribes	0.0211	0.0237	0.0251	0.0254
	(0.0550)	(0.0544)	(0.0545)	(0.0547)

NOTES: This table presents additional robustness checks controlling for pre-villagization political turnout at elections and support for Nyerere's TANU Party. Results are presented for the coefficients on *young x treatment*, which provides the difference-in-difference estimate comparing the mean outcome variable for older cohorts born prior to 1960 to a younger cohort born after 1970. Column (1) repeats estimates using a full set of controls as in column (6) of tables 5-8, chapter 2. Outcome variables are obtained from various rounds of the Tanzania Afrobarometer Survey. Additional controls on pre-villagization election results are obtained from results compiled by the Election Study Committee (1974). Robust standard errors are in parentheses and clustered at the 1978 district boundaries. *** $p<0.01$, ** $p<0.05$, * $p<0.1$.

References

ACEMOGLU, D., S. JOHNSON, AND J. ROBINSON (2001): "The colonial origins of comparative development: an empirical investigation," *American Economic Review*, 91, 1369–401.

——— (2002): "Reversal of fortune: geography and institutions in the making of the modern world income distribution," *Quarterly Journal of Economics*, 117, 1231–94.

ACHEN, C. (1992): "Social psychology, demographic variables, and linear regression: Breaking the iron triangle in voting research," *Political Behavior*, 14(3), 195–211.

AGADZI, V., Y. ABOAGYE-ATTA, J. NELSON, D. HOPKINS, AND P. PERINE (1985): "Yaws in Ghana," *Review of Infectious Diseases*, 7, Supplement 2, International Symposium on Yaws and Other Endemic Treponematoses, S233–S236.

ALESINA, A., R. BAQIR, AND W. EASTERLY (1999): "Public Goods and Ethnic Divisions," *Quarterly Journal of Economics*, 114 (1999), 1243–1284.

ALESINA, A., AND E. L. FERRARA (2000): "Participation in Heterogeneous Communities," *Quarterly Journal of Economics*, 115 (2000), 847–904.

ALESINA, A., AND N. FUCHS-SCHUENDELN (2007): "Good-bye Lenin (or Not?): The Effect of Communism on People's Preferences," *American Economic Review*, 97 No. 4.

ALESINA, A., P. GIULIANO, AND N. NUNN (2011): "On the Origins of Gender Roles: Women and the Plough," *NBER Working Paper No. 17098*.

ANGRIST, J., AND J. PISCHKE (2008): *Mostly harmless econometrics: an empiricist's companion*. Princeton: Princeton University Press.

ANSOLABEHERE, S., AND J. SNYDER (2006): "Party Control of State Government and the Distribution of Public Expenditures," *Scandinavian Journal of Economics*, 108, no. 4, 547–569.

AUSTEN, R. A. (1987): *African Economic History: Internal Development and External Dependency*. London: James Currey.

AUSTIN, D. (1964): *Politics in Ghana, 1946-1960*. London: Oxford University Press.

BANERJEE, A., AND L. IYER (2005): "History, institutions and economic performance: the legacy of colonial land tenure systems in India," *American Economic Review*, 95, 1190–213.

BANFUL, A. B. (2009): "Do Institutions Limit Clientelism? A Study of the District Assemblies Common Fund in Ghana," *IFPRI Discussion Paper*, 00855.

BANK, W. (1961): *The Economic Development of Tanganyika*. Baltimore: Johns Hopkins University Press.

BARKAN, J. D. (1984): *Politics and Public Policy in Kenya and Tanzania*. New York: Praeger Publishers.

BARKER, J. S., AND J. SAUL (1974): *The Tanzania Elections in Post-Arusha Perspective*. Dar es Salaam: Tanzania Publishing House Ltd.

BAUER, P. (2005): *Afrobarometer Data Codebook for Ghana (Round 2)*. mimeo Michigan State University.

BESLEY, T., AND T. PERSSON (2009): "The Origins of State Capacity: Property Rights, Taxation and Politics," *American Economic Review*, 99, 1218–1244.

——— (2011a): "Fragile States and Development Policy," *LSE mimeo*.

——— (2011b): *Pillars of Prosperity*. Princeton University Press.

BLEAKLEY, H. (2007): "Disease and Development: Evidence from Hookworm Eradication in the American South," *Quarterly Journal of Economics*, February 2007.

——— (2010): "Malaria eradication in the Americas: A retrospective analysis of childhood exposure," *American Economic Journal: Applied Economics*, 2(2), 1–45.

BOATENG, E., T. AYAMGA, T. CHIDLEY, J. ELGY, D. SAVORY, AND H. VAN VELTHUIZEN (1999): "Land Suitability Analysis - Geo-spatial modelling for agro-ecological assessments and decision support for development planning," *GERMP and EISD Project Report, Soil Research Institute of Ghana*.

BOCKSTETTE, V., A. CHANDA, AND L. PUTTERMAN (2002): "States and markets: the advantage of an early start," *Journal of Economic Growth*, 7, 347–69.

BOMMIER, A., AND S. LAMBERT (1999): "Education Demand and Age at School Enrollment in Tanzania," *Journal of Human Resources*, XXXV(1).

BOUND, J., D. A. JAEGER, AND R. M. BAKER (1995): "Problems with Instrumental Variables Estimation When the Correlation between the Instruments and the Endogenous Explanatory Variable Is Weak," *Journal of the American Statistical Association*, 90 (June), 443–50.

CAMPBELL, A., P. CONVERSE, W. MILLER, AND D. STOKES (1960): *The American Voter*. New York: John Wiley and Sons.

CANNING, D. (2006): "Priority setting and the "neglected" tropical diseases," *Trans. of the Royal Society of Tropical Medicine and Hygiene*, 100, 499–504.

CARDINALL, A. W. (1932): *The Gold Coast, 1931 : A review of conditions in the Gold coast in 1931 as compared with those of 1921, based on figures and facts collected by the chief census officer of 1931, together with a historical, ethnographical and sociological survey of the people of that country*. Accra: Gold Coast Government Printer.

CHALIGHA, A., R. MATTES, M. BRATTON, AND Y. D. DAVIDS (2002): "Uncritical Citizens or Patient Trustees? Tanzanians' Views of Political and Economic Reform," *Afrobarometer Working Paper*, 18.

CHAYANOV, A. V. (1966): *Theory of the Peasant Economy*. Homewood, Ill., Published for the American Economic Association.

CHENERY, H. (1971): *Studies in Development Planning*. Harvard University Press.

CHENERY, H. B. (1975): "The Structuralist Approach to Development Policy," *American Economic Review*, 65, No. 2, 310–316.

CHRISTIANSEN, S. (1963): "Report on a visit to Ghana (2-25 October 1962)," *World Health Organization*, AFR/VDT/23.

CLIFFE, L. (1967): *One Party Democracy: The 1965 Tanzania General Elections*. Nairobi: East African Publishing House.

COCOA CHOCOLATE AND CONFECTIONERY ALLIANCE (1951): "Report of the Cocoa Conference held at Grosvenor House, London, W.1., 11th to 13th September, 1951," .

——— (1957): "Report of the Cocoa Conference held at Grosvenor House, London, W. 1, 10th to 12th September, 1957," .

COHEN, D., R. E. NISBETT, B. F. BOWDLE, AND N. SCHWARZ (1996): "Insult, aggression, and the southern culture of honor: An "experimental ethnography"," *Journal of Personality and Social Psychology*, 70(5), 945–960.

COLDHAM, S. (1995): "Land Tenure Reform in Tanzania: Legal Problems and Perspectives," *Journal of Modern African Studies*, 33, 2 (1995), 227–242.

COLLIER, P., R. RADWAN, AND S. WANGWE (1986): *Labour and Poverty in Rural Tanzania: Ujamaa and Rural Development in the United Republic of Tanzania*. Oxford: Clarendon Press.

COLONIAL OFFICE (1948): *Report of the Commission of Enquiry into Disturbances in the Gold Coast (Led by Aiken Watson)*. London: His Majesty's Stationery Office, Colonial No. 231.

CONLEY, T., C. HANSEN, AND P. E. ROSSI (2012): "Plausibly Exogenous," *Review of Economics and Statistics*, 94(1), 260–272.

CONTEH, L., T. ENGELS, AND D. MOLYNEUX (2010): "Socioeconomic aspects of neglected tropical diseases," *Lancet*, 375, 239–247.

CONVERSE, P. E. (1969): "Of time and partisan stability," *Comparative Political Studies*, 2 (July), 139–71.

CORNWELL, P. B. (1958): "Movements of the Vectors of Virus Diseases of Cacao in Ghana I: Canopy Movement in and between Trees," *Bulletin of Entomological Research*, 49(3) Sep 1958.

——— (1960): "Movements of the vectors of virus diseases of cacao in Ghana II: Wind movements and aerial dispersal," *Bulletin of Entomological Research*, 51(1) Apr 1960.

COULSON, A. (1982): *Tanzania: A Political Economy*. New York: Oxford University Press.

COX, G. W., AND M. D. MCCUBBINS (1986): "Electoral Politics as a Redistributive Game," *Journal of Politics*, 48, 370–389.

CUTLER, D., W. FUNG, M. KREMER, M. SINGHAL, AND T. VOGL (2010): "Early-life malaria exposure and adult outcomes: evidence from malaria eradication in India," *American Economic Journal: Applied Economics*, 2(2), 72–94.

DAHLBERG, M., AND E. JOHANNSON (2002): "On the Vote Purchasing Behavior of Incumbent Governments," *American Political Science Review*, 96, 27–40.

DANQUAH, F. K. (2003): "Sustaining a West African Cocoa Economy: Agricultural Science and the Swollen Shoot Contagion in Ghana, 1936-1965," *African Economic History*, 31, 43–74.

DE VRIES, J. (1961): "The role of schools in yaws surveillance," *mimeo World Health Organization, SEA/VDT/Yaws Conf. 10*.

DE VRIES, J., AND L. P. FORTMANN (1979): *Large scale villagization: Operation Sogeza in Iringa*. Nottingham: Spokesman.

DELL, M. (2010): "The Persistent Effects of Peru's Mining Mita," *Econometrica*, 78(6), 2010, 1863–1903.

——— (2011): "Insurgency, agrarian reform, and long-run development: Lessons from the Mexican Revolution," *mimeo Massachusetts Institute of Technology*.

DICKSON, K. B. (1969): *A historical geography of Ghana*. London: Cambridge University Press.

DIPPEL, C. (2011): "Forced Coexistence and Economic Development: Evidence from Native American Reservations," *mimeo, UCLA Anderson School of Management*.

DIXIT, A., AND J. LONDREGAN (1995): "Redistributive Politics and Economic Efficiency," *American Political Science Review*, 89, 856–866.

——— (1996): "The Determinants of Success of Special Interests in Redistributive Politics," *Journal of Politics*, 58, 1132–1155.

DUFLO, E. (2001): "Schooling and Labor Market Consequences of School Construction in Indonesia: Evidence from an Unusual Policy Experiment," *American Economic Review*, 91(4), 795–813.

EAMA (1962): *Monthly and Annual Rainfall in Tanganyika and Zanzibar during the 30 years, 1931-1960*.

EDWARDS, S. (2012): "Is Tanzania a Success Story? A Long Term Analysis," *NBER Working Paper 17764*.

ELECTION STUDY COMMITTEE (1974): *Socialism and participation: Tanzanias 1970 National elections*. Dar es Salaam: Tanzania Pub. House.

ELLIS, F. (1982): "Agricultural Price Policy in Tanzania," *World Development*, 10, No.4, 263–283.

ENGERMAN, S., AND K. L. SOKOLOFF (1997): *Factor endowments, institutions, and differential paths of growth among New World economies: a view from economic historians of the United States. In How Latin America Fell Behind*. Stanford, CA: Stanford Univ. Press.

——— (2002): "Factor endowments, inequality, and paths of development among New World economies," *NBER Working Paper 9259*.

ERGAS, Z. (1980): "Why Did the Ujamaa Village Policy Fail? — Towards a Global Analysis," *Journal of Modern African Studies*, 18, No. 3 (Sep., 1980), 387–410.

FASQUELLE, A. (1973): "Treponematoses control in Ghana, Report on a mission, 27 April -12 June 1973," *World Health Organization, AFR/VDT/32*.

FEINSTEIN, C. H. (2005): *An Economic History of South Africa*. Cambridge: Cambridge University Press.

FENSKE, J. (2010): "The Causal History of Africa: A Response to Hopkins," *Economic History of Developing Regions*, 25, Issue 2, 2010.

FIORINA, M. P. (1981): *Retrospective Voting in American National Elections*. New Haven, CT: Yale University Press.

FORTMANN, L. (1980): *Peasants, Officials and Participation in Rural Tanzania: Experience with Villagization and Decentralization. Special series on Rural Local Organization*. Ithaca NY: Cornell University.

FREUND, W. M. (1981): "Class Conflict, Political Economy and the Struggle for Socialism in Tanzania," *African Affairs*, 80, No. 321 (Oct., 1981), 483–499.

GELLNER, E., AND C. MICAUD (1972): *Arabs and Berbers: From Tribe to Nation in North Africa*. Lexington, Mass.: D.C. Heath and Co.

GENNAIOLI, N., AND I. RAINER (2007): "The modern impact of precolonial centralization in Africa," *Journal of Economic Growth*, 12, 185–234.

GERBER, A. S., G. A. HUBER, AND E. WASHINGTON (2010): "Partisan Affiliation, Partisanship and Political Beliefs: A Field Experiment," *American Political Science Review*, 104 (4), 720–744.

GIULIANO, P., AND A. SPILIMBERGO (2009): "Growing Up in a Recession: Beliefs and the Macroeconomy," *NBER Working Paper 15321*.

GOLD COAST COLONY (1935): *Report on the Medical Department for the year 1933-34*. London: Crown Agents for the Colonies.

GREEN, D., B. PALMQUIST, AND E. SCHICKLER (2002): *Partisan Hearts and Minds: Political Parties and the Social Identities of Voters*. New Haven: Yale University Press.

GREIF, A. (1994): "Cultural beliefs and the organization of society: A historical and theoretical reflection on collectivist and individualist societies," *Journal of Political Economy*, 102(5), 912–950.

GUISO, L., P. SAPIENZA, AND L. ZINGALES (2007): "Social Capital as Good Culture," *NBER Working Paper 13712*.

——— (2008): "Long-Term Persistence," *NBER Working Paper 14278*.

HACKETT, C. J. (1953): "Extent and Nature of the Yaws Problem in Africa," *Bulletin of the World Health Organization*, 8, 299–343.

HERBST, J. (2000): *States and Power in Africa: Comparative lessons in authority and control*. Princeton: Princeton University Press.

HERRICK, A., S. HARRISON, H. JOHN, S. MACKNIGHT, AND B. SKAPA (1968): *Area Handbook for Tanzania*. Washington DC: US Government Printing Office.

HILL, P. (1956): *The Gold Coast Cocoa Farmer*. London: Oxford University Press.

HOLLINGER, S., AND S. CHANGNON (1993): "Response of corn and soybean yields to precipitation augmentation and implications for weather modification in Illinois," *Bulletin (Illinois State Water Survey)*, 73.

HOPKINS, A. (2009): "The New Economic History of Africa," *Journal of African History*, 50 (2009), 155–77.

HOPKINS, A. G. (1973): *An Economic History of West Africa*. London: Longmans.

HOTEZ, P., A. FENWICK, L. SAVIOLI, AND D. MOLYNEUX (2009): "Rescuing the bottom billion through control of neglected tropical diseases," *Lancet*, 373, 1570–75.

HUILLERY, E. (2009): "History Matters: The Long Term Impact of Colonial Public Investments in French West Africa," *American Economic Journal — Applied Economics*, 1(2), 176–215.

HYDEN, G. (1980): *Beyond Ujamaa in Tanzania: Underdevelopment and an Uncaptured Peasantry*. London: Heinemann Educational Books.

ILIFFE, J. (1975): *Tanzania in Maps*. London: University of London Press.

——— (1979): *A Modern History of Tanganyika*. Cambridge: Cambridge University Press.

JACOB, M., AND M. TYRELL (2010): "The Legacy of Surveillance: An Explanation for Social Capital Erosion and the Persistent Economic Disparity between East and West Germany," *Unpublished working paper, EBS Universität für Wirtschaft und Recht, Wiesbaden and Oestrich-Winkel, Germany; and Zeppelin University, Friedrich-Shafen, Germany*.

JENSEN, S., AND J. MKAMA (1968): *District Data, Tanzania 1967*. Dar es Salaam: Ministry of Economic Affairs and Development Planning.

JOHNSON, D. H. (1962): "Rain in East Africa," *Quarterly Journal of the Royal Meteorological Society*, 88, No. 375.

KAPLAN, B. (1961): "New settlement and agricultural development in Tanganyika," *mimeo (unpublished paper) Dar es Salaam*.

KEENY, S. M., AND M. GAAN (1961): "The role of UNICEF in yaws control and its post-campaign activities," *Report of the Inter-regional Yaws Conference and Coordination Meeting, Bandung, Indonesia, 30 October - 10 November, 1961, mimeo UNICEF/WHO*.

KHEMANI, S. (2007): "Does delegation of fiscal policy to an independent agency make a difference? Evidence from intergovernmental transfers in India," *Journal of Development Economics*, 82, Issue 2, March 2007, 464–484.

KILLICK, T. (1976): "The possibilities of development planning," *Oxford Economic Papers*, 28(2), 161–184.

——— (1983): "Development Planning in Africa: Experiences, Weaknesses and Prescriptions," *Development Policy Review*, 1(1) May 1983, 47–76.

KOFF, D., AND G. VON DER MUHLL (1967): "Political Socialisation in Kenya and Tanzania — A Comparative Analysis," *Journal of Modern African Studies*, 5, No. 1 (May, 1967), 13–51.

LASSWELL, H. (1936): *Politics: Who Gets What, When, How*. New York, London: McGraw-Hill.

LIESE, B., AND L. SCHUBERT (2009): "Official development assistance for health - how neglected are neglected tropical diseases? An analysis of health financing," *International Health*, 1, 141–47.

LINDBECK, A., AND J. W. WEIBULL (1987): "Balanced-Budget Redistribution as the Outcome of Political Competition," *Public Choice*, 52, 273–297.

LOFCHIE, M. F. (1978): "Agrarian Crisis and Economic Liberalisation in Tanzania," *Journal of Modern African Studies*, 16, No. 3 (Sep., 1978), 451–475.

LUCAS, A. M. (2010): "Malaria eradication and educational attainment: evidence from Paraguay and Sri Lanka," *American Economic Journal: Applied Economics*, 2, 46–71.

MADESTAM, A., D. SHOAG, S. VEUGER, AND D. YANAGIZAWA-DROTT (2011): "Do political protests matter? Evidence from the Tea Party Movement," *mimeo Harvrd Kennedy School*.

MADESTAM, A., AND D. YANAGIZAWA-DROTT (2011): "Shaping the Nation: The Effect of Fourth of July on Political Preferences and Behavior in the United States," *mimeo Harvard Kennedy School*.

MAMDANI, M. (1996): *Citizen and Subject: Contemporary Africa and the Legacy of Late Colonialism*. Princeton: Princeton University Press.

MARO, P. S., AND W. MLAY (1982): *Population Redistribution in Tanzania (in Redistribution of population in Africa)*. London; Exeter, N.H. : Heinemann Educational, 1982.

MAURO, P. (1995): "Corruption and Growth," *Quarterly Journal of Economics*, CX (1995), 681–712.

MBILINYI, M. J. (1976): "Peasants' Education in Tanzania," *The African Review*, 6 (2).

MCCALL, M. (1985): *Environmental and agricultural impacts of Tanzanias villagization programme, in Population and Development Projects in Africa*. Cambridge [Cambridgeshire]; New York : Cambridge University Press.

MCHENRY, D. E. (1979): *Tanzania's Ujamaa villages: The Implementation of a Rural Development Strategy*. Berkeley: Institute of International Studies.

MCHENRY, D. E. (1981): *Ujamaa villages in Tanzania: a bibliography*. Uppsala: Scandinavian Institute of African Studies.

MCINTOSH, S. K. (1999): *Beyond chiefdoms: pathways to complexity in Africa*. Cambridge, U.K.: Cambridge University Press.

MICHALOPOULOS, S., AND E. PAPAIOANNOU (2011): "Divide and Rule or the Rule of the Divided? Evidence from Africa," *NBER Working Paper 17184*.

MIGUEL, E. (2004): "Tribe or Nation? Nation Building and Public Goods in Kenya versus Tanzania," *World Politics*, 56, no. 3, 327–362.

MIGUEL, E., AND M. K. GUGERTY (2005): "Ethnic Diversity, Social Sanctions, and Public Goods in Kenya," *Journal of Public Economics*, 89 (2005), 2325–2368.

MIGUEL, E., AND M. KREMER (2004): "Worms: identifying impacts on education and health in the presence of treatment externalities," *Econometrica*, LXXII, 159–217.

MOLYNEUX, D. (2008): "Combating the "other diseases" of MDG 6: changing the paradigm to achieve equity and poverty reduction?," *Trans of the Royal Society of Tropical Medicine and Hygiene*, 102, 509–19.

MOORE, J. E. (1973): *Population distribution and density*. Dar es Salaam: BRALUP (The Population of Tanzania, Census Volume 6).

MUZZINI, E., AND W. LINDEBOOM (2008): *The Urban Transition in Tanzania: Building the Empirical Base for Policy Dialogue*. Washington DC: The World Bank.

NARAYAN, D., AND L. PRITCHETT (1999): "Cents and Sociability: Household Incomes and Social Capital in Rural Tanzania," *Economic Development and Cultural Change*, 47:4 (July 1999).

NGWARE, S., AND M. HAULE (1992): *The Forgotten Level: Village Government in Tanzania*. Hamburg: Institut fuer Afrika-Kunde.

NIEMI, R. G., AND M. K. JENNINGS (1991): "Issues and Inheritance in the Formation of Party Identification," *American Journal of Political Science*, 35, 970–988.

NIEUWOLT, S. (1977): "The influence of rainfall on rural population distribution in Tanzania," *Journal of Tropical Geography*, 44-47 (1977-78).

NORD, R., Y. SOBOLEV, D. DUNN, A. N. H. ALEJANDRO HAJDENBERG, S. MAZIAD, AND S. ROUDET (2009): *Tanzania: the story of an African transition*. Washington, D.C.: International Monetary Fund, Africa Department.

NORTH, D. C., AND R. P. THOMAS (1973): *The Rise of the Western World*. Cambridge, UK: Cambridge Univ. Press.

NUNN, N. (2009): "The importance of history for economic development," *Annual Review of Economics*, 1(1), 65–92.

——— (2012): "Culture and the Historical Process," *Economic History of Developing Regions*, 27, 2012, S108–S126.

NUNN, N., AND L. WANTCHEKON (2011): "The Slave Trade and the Origins of Mistrust in Africa," *American Economic Review*, Vol. 101, No. 7, December 2011, 3221–3252.

NYERERE, J. (1966): *Freedom and unity: A selection from writings and speeches, 1952-65*. London: Oxford University Press.

——— (1967): *Education for Self-Reliance*. Dar es Salaam: Government Printers.

——— (1977): *Arusha Declaration: Ten Years After*. Dar es Salaam: Government Printers.

OMARI, C. (1984): *Towards rural development in Tanzania: Some issues on policy implementation in the 1970s*. Arusha: Eastern Africa Publications Limited.

ONORI, E. (1962): "The results of a mass campaign against yaws in the Volta Region of Ghana," *Ghana Medical Journal*, December 1962.

OSILI, U. O., AND B. T. LONG (2008): "Does female schooling reduce fertility? Evidence from Nigeria," *Journal of Development Economics*, 87 (2008), 57–75.

PAVEL, M. (1963): "The use of statistics in health planning: achievement of balanced services," *Internal Report - Division of Health Statistics, Ministry of Health, Ghana*.

PERINE, P. L., D. R. HOPKINS, P. NIEMEL, G. CAUSSE, G. ANTAL, AND R. S. JOHN (1984): "Handbook of endemic treponematoses: yaws, endemic syphilis and pinta," *Geneva: The World Health Organization*.

POPKIN, S. (1991): *The reasoning voter: Communication and persuasion in presidential campaigns*. Chicago: University of Chicago Press.

PORTA, R. L., F. LOPEZ DE SILANES, A. SHLEIFER, AND R. VISHNY (1997): "Legal determinants of external finance," *Journal of Finance*, 52, 1131–50.

——— (1998): "Law and finance," *Journal of Political Economy*, 106, 1113–55.

POSNER, D. N. (2004): "Measuring Ethnic Fractionalization in Africa," *American Journal of Political Science*, 48, No. 4, October 2004, 849–863.

PREWITT, K. (1971): *Education and Political Values: An East African Case Study*. Nairobi: East African Publishing House.

PREWITT, K. G., G. V. DER MUHLL, AND D. COURT (1970): "School Experiences and Political Socialization: A Study of Tanzanian Secondary School Students," *Comparative Political Studies*, 3, no. 2 (1970).

PUTNAM, R. D. (1993): *Making Democracy Work*. Princeton: Princeton University Press.

PUTTERMAN, L. (1981): "Is a Democratic Collective Agriculture Possible? Theoretical Considerations and Evidence from Tanzania," *Journal of Development Economics*, 9, 375–403.

——— (1986): *Peasants, Collectives, and Choice: Economic Theory and Tanzania's Villages*. Greenwich CT: Jai Press Inc.

——— (1995a): "Economic Reform and Smallholder Agriculture in Tanzania: A Discussion of Recent Market Liberalization, Road Rehabilitation, and Technology Dissemination Efforts," *World Development*, 23, 311–26.

——— (1995b): "Social Capital and Development Capacity: The Example of Rural Tanzania," *Development Policy Review*, 13, 5–22.

RAIKES, P. L. (1975): "Ujamaa and Rural Socialism," *Review of African Political Economy*, 3 (May - Oct., 1975), 33–52.

——— (1978): "Rural differentiation and class formation in Tanzania," *Journal of Peasant Studies*, 3, 285–325.

——— (1986): *Eating the carrot and wielding the stick: The agricultural sector in Tanzania*. Scandinavian Institute of African Studies.

ROSEI, L. (1963): "Control of endemic yaws in the Brong Ahafo region of Ghana: A report on a mass treatment campaign," *West African Medical Journal*, February 1963.

ROSS, S., AND J. BROATCH (1951): "A review of the swollen shoot control campaign in the Gold Coast," *Report of the Cocoa Conference held at Grosvenor House, London 11-13 September 1951*.

RWEYEMAMU, A. H., AND B. MWANSASU (1974): *Planning in Tanzania: background to decentralisation*. Nairobi: East African Literature Bureau.

SALAMAN, S. (1980): "Ethnic differences in farm family land transfers," *Rural Sociology*, 45(2), 290–308.

SAMOFF, J. (1981): "Crises and Socialism in Tanzania," *Journal of Modern African Studies*, 19, No. 2 (Jun., 1981), 279–306.

SCOTT, D. (1959): "The first stage of a campaign against yaws in Northern Ghana 1956-1959: The assessment of a joint WHO/UNICEF/Ghana endeavour," *mimeo World Health Organization, INT/VDT/145*.

SCOTT, J. C. (1999): *Seeing Like a State: How Certain Schemes to Improve the Human Condition have Failed*. New Haven: Yale University Press.

SCOTT, J. C. (2009): *The Art of Not Being Governed: An Anarchist History of Upland Southeast Asia*. New Haven: Yale University Press.

SETTLE, J. E., C. T. DAWES, AND J. H. FOWLER (2009): "The Heritability of Partisan Attachment," *Political Research Quarterly*, 62 Number 3, September 2009, 601–613.

SHEFFIELD, J. R. (1979): "Basic Education for the Rural Poor: The Tanzanian Case," *Journal of Developing Areas*, 14, No. 1 (Oct., 1979).

SHIVJI, I. G. (1976): *Class Struggles in Tanzania*. London: Heinemann.

SHIVJI, I. G. (1994): *Report of the Presidential Commission of Inquiry into Land Matters, Vol. I, Land Policy and Land Tenure Structure (Chaired by Prof. I. G. Shivji)*. Ministry of Lands, Housing, and Urban Development, Tanzania, in co-operation with the Scandinavian Institute of African Studies, Uppsala, Sweden.

SHIVJI, I. G., AND C. M. PETERS (2003): *Village Democracy Initiative*. Report Prepared for The President's Office - Regional Administration and Local Government.

SKARSTEIN, R. (2005): "Economic Liberalization and Smallholder Productivity in Tanzania. From Promised Success to Real Failure, 1985-1998," *Journal of Agrarian Change*, 5(3).

SOETOPO, M., AND R. WASITO (1953): "Experience with yaws control in Indonesia: Preliminary results with a simplified approach," *Bulletin of the World Health Organization*, 8(1-3), 273–291.

STABLER, E. (1979): “Kenya and Tanzani: Strategies and Realities in Education and Development,” *African Affairs (London)*, 78(310), 33–56.

STAIGER, D., AND J. H. STOCK (1997): “Instrumental Variables Regression with Weak Instruments,” *Econometrica*, 65 (May), 557–86.

STOCK, J. H., J. H. WRIGHT, AND M. YOGO (2002): “A Survey of Weak Instruments and Weak Identification in Generalized Method of Moments,” *Journal of Business and Economic Statistics*, 20, 518–29.

TABELLINI, G. (2008): “Culture and institutions: Economic development in the regions of Europe,” *Journal of the European Economic Association*, 8(4), 677–716.

TANBURN, E. (1957): “The long-term prospects of Ghana Cocoa Production,” *mimeo, Ghana Ministry of Food and Agriculture*.

TANZANIA (1968): *1967 Tanzania Population Census*. Dar es Salaam: United Republic of Tanzania, Central Census Office.

——— (1979): *1978 Population Census*. Dar es Salaam : United Republic of Tanzania, Bureau of Statistics, Ministry of Planning and Economic Affairs.

THOMAS, I. (1982): *Population data and development planning in Tanzania: Tanzania 1978 Population Census User Notes*. Norwich UK: Overseas Development Group.

TIME MAGAZINE (1955): “Africa v. Yaws,” *Time*, 66, Issue 26.

VAN ARDAKIE, B. (1972): *Planning in Tanzania in* The Crisis in Planning. London: Chatto and Windus.

VERHAGEN, K. (1980): “Tanzanian Rural Development Policy, 1975-78,” *Development and Change*, 11(1980), 285–95.

VON DER MUHLL, G. (1971): *Education, Citizenship and Social Revolution in Tanzania, in Education and Political Values: An East African Case Study*. Nairobi: East African Publishing House.

VON FREYHOLD, M. (1979): *Ujamaa Villages in Tanzania: Analysis of a Social Experiment*. London: Heinemann.

WADDY, B. B. (1956): "Organisation and work of the Gold Coast medical field units," *Trans. Royal Society of Tropical Medicine and Hygiene*, 50, 313–343.

WADE, R. (1988): *Village Republics*. Cambridge: Cambridge University Press.

WEBER, E. (1976): *Peasants into Frenchmen: The Modernization of Rural France, 1870-1914*. Stanford: Stanford University Press.

ZAHRA, A. (1956): "Yaws eradication campaign in Nsukka Division, Eastern Nigeria," *Bulletin of the World Health Organization*, 15, 911–935.

ZELEZA, P. T. (1993): *A Modern Economic History of Africa, vol. I: The Nineteenth Century*. Dakar: CODESRIA.